

AD-A053 633

MARYLAND UNIV BALTIMORE COUNTY BALTIMORE DIV OF MATH--ETC F/G 12/1  
ESTIMATION AND DETECTION OF LIE GROUPS.(U)

OCT 77 J T LO

AFOSR-74-2671

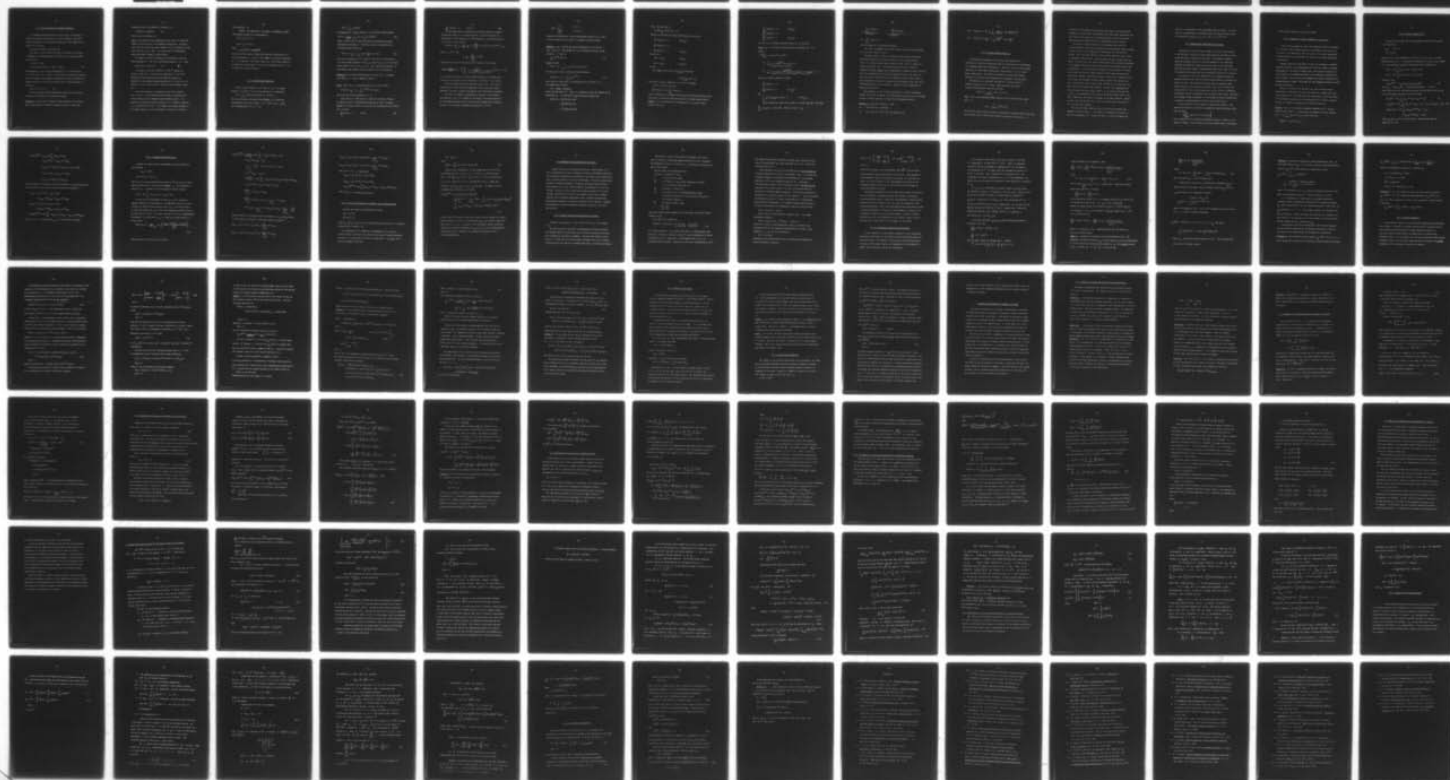
UNCLASSIFIED

MRR-77-11

AFOSR-TR-78-0734

NL

1 OF 2  
ADA  
053633





# Mathematics Research Reports

AD A 053633



AD No. ~~1~~  
DDC FILE COPY

## DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited

Division of Mathematics and Physics

UNIVERSITY OF MARYLAND  
BALTIMORE COUNTY

Baltimore, Maryland 21228

Q/D





ARMY STAFF OFFICE OF SCIENTIFIC RESEARCH (ASOR)  
SECTION OF SUBMITTAL TO DOD  
This technical report has been reviewed and is  
approved for public release IAW ARS 160-20 (70).  
Distribution is unlimited.  
A. B. HARRIS  
Technical Information Officer

AD A 053633

Estimation And Detection On Lie Groups

by

James Ting-Ho Lo

Mathematics Research Report No. 77-11

October 1977

AD No.   
DDC FILE COPY

DDC  
RECEIVED  
MAY 8 1978  
B

ACKNOWLEDGMENT

This work was supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-74-2671B.

DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited



# ESTIMATION AND DETECTION ON LIE GROUPS

James Ting-Ho Lo  
Department of Mathematics  
University of Maryland Baltimore County  
Baltimore, Maryland 21228

## CONTENTS

### I. INTRODUCTION

### II. PROBABILITY ON THE CIRCLE

1. The Distribution Function
2. The Characteristic Function
3. Error Criteria and Optimal Estimates
4. Folded Normal Densities
5. Circular Normal Densities

### III. DISCRETE-TIME ESTIMATION ON THE CIRCLE

1. Exponential Fourier Densities on the Circle
2. A Basic System on the Circle
3. A Phase-Shift-Keyed System
4. Periodic Measurements in Additive White Gaussian Noise

### IV. CONTINUOUS-TIME ESTIMATION ON THE CIRCLE

1. Signal Processes and Observation Processes
2. Conditional Probability Distributions
3. Optimal Estimation
4. Random Initial State
5. Multichannel Estimation

ACCESSION for		
NTIS	White Section	<input checked="" type="checkbox"/>
DDC	Buff Section	<input type="checkbox"/>
UNANNOUNCED		<input type="checkbox"/>
JUSTIFICATION _____		
BY _____		
DISTRIBUTION/AVAILABILITY CODES		
Dist.	AVAIL	and/or SPECIAL
A		



## I. INTRODUCTION.

In the past, most detection, estimation, and control problems were studied in a linear space setting. While the linear space approach leads to simple solutions for linear systems, no effective synthesis procedures for optimal detection, estimation, and control have been obtained for large classes of nonlinear systems.

It is only natural to believe that a nonlinear problem is best studied in some kind of a nonlinear space. Among all possible nonlinear spaces it is only natural to start with a space which is locally linear, on which a differential calculus can be used, and which has a group structure for us to utilize. Such a nonlinear space does exist in mathematical literature and is called a Lie group. In fact, it was invented by Sophus Lie to study nonlinear differential equations. The theory of Lie groups has been well established and provides us with a large chest of geometric and algebraic tools.

In addition to the mathematical nicety, the Lie groups are natural state spaces for many nonlinear problems of practical importance. Notable examples are the rotation groups, which are the state spaces for frequency demodulation, gyroscopic analysis, and satellite attitude estimation and control. Other examples can be found in power conversion, nuclear reactor control, and compartmental-model study in bioscience, etc.

Recent years have seen many useful and interesting results on detection, estimation, and control problems with Lie group structures.



Most of these results are facilitated by the rich geometric and algebraic structures which are inherent in these problems and are made clear only in a Lie group setting. The reader is referred to [1]-[3], from which most related articles that are not in the reference list of this chapter can be traced. This chapter is not intended to be a survey of the development of what is now called the geometric approach. We will rather restrict our attention mainly to estimation and detection and some closely related issues.

In contrast to the linear theory, the continuous-time and the discrete-time systems on Lie groups are very different in nature. The approaches to their estimation and detection problems are thus very different and have been developed on the bases of two separate ideas. The idea for continuous-time systems is that of "rolling without slipping." The idea for discrete-time systems is the use of the exponential Fourier densities.

The continuous-time systems on a Lie group that correspond to the linear systems on a linear space are bilinear in form. In fact, the bilinear systems can be viewed as induced by the linear systems through "rolling without slipping." Furthermore, "rolling without slipping" can be shown to be an "almost sure" bijective mapping between the bilinear systems and the linear systems. It is known that the local study of a Lie group is entirely equivalent to the study of the finite-dimensional linear algebraic structures of the associated Lie algebra. "Rolling without slipping" does indeed facilitate similar simplification in studying estimation and detection.

The exponential Fourier densities have been used to derive finite-dimensional optimal estimation schemes for many discrete-time systems on compact Lie groups. This is made possible mainly by the closure property of the exponential Fourier densities of any given finite order under the operation of taking conditional distributions. Another reason for using exponential Fourier densities is that any continuous or bounded-variation probability density on a compact Lie group can be approximated as closely as desired by such a density.

Most of these ideas can be clearly illustrated on the unit circle, the simplest compact Lie group. The circle is also the natural state space for many estimation and detection problems of practical importance such as frequency and phase demodulation and single-degree-of-freedom gyroscopic analysis. Therefore a detailed theory of estimation on the circle will be presented in the next three sections. No knowledge of Lie groups is required to understand them. Estimation and detection on general Lie groups are studied in the last two sections. The required definitions and theorems from the Lie theory are briefly summarized there.

Although the two sections on general Lie groups and the three sections on the circle can be read independently, an understanding of the circle case can definitely help understand the problems and the results on general Lie group. The main references for Sections II-VI are [4]-[8] respectively. Section V is the only section that contains some new results.

This chapter is not intended to exhaust all existing results on estimation and detection problems with Lie group structure. The interested reader is referred to [9]-[17] for some of these results beyond this chapter.

## II. PROBABILITY ON THE CIRCLE.

There are many fundamental differences between the estimation and detection problems on Euclidean spaces and those on Lie groups. In order for some readers to appreciate them, this section will be addressed to some probabilistic elements on the circle. The probability distribution function and the characteristic function on the circle will first be briefly introduced.

One of the main concerns in this chapter is to study how one uses the knowledge of the probability distribution of a random variable taking values on a Lie group to determine an estimate of the random variable that minimizes a certain error criterion. The conventional least squares technique cannot be used here. Let us take the circle as an example. The square error of the angles  $0^\circ$  and  $359^\circ$  is  $(359^2)^\circ$ , whereas by geometrical intuition they are only  $1^\circ$  apart. In Subsection II.3 we will look into this issue on the circle in detail.

The importance of the normal probability densities cannot be overemphasized for estimation and detection on Euclidian spaces. Unfortunately, there does not exist an analogous density on the



circle that possesses all the nice properties of the normal density. In fact, the nice properties of the normal density are almost equally divided between two contenders for normalcy, the folded normal density and the circular normal density. It turns out that while the folded normal density is natural to use for continuous-time estimation, the circular normal density is more suitable for discrete-time estimation. They will both be discussed and compared in this section.

## II. 1. The Distribution Function.

A point on the unit circle  $S^1$  can be represented by either the angle  $\theta \in (-\pi, \pi)$  it makes with a fixed reference point on the circle or by the  $2 \times 2$  orthogonal matrix

$$\exp R\theta = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \approx I + \theta R$$

where the matrix  $R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$  is called the infinitesimal rotation matrix. The addition of two angles  $\theta_1$  and  $\theta_2$  modulo  $2\pi$ , denoted as  $\theta_1 \oplus \theta_2$ , corresponds to the multiplication of the two matrices representing the points.

Let  $\theta$  be a random variable taking values on  $S^1 = (-\pi, \pi]$ . The distribution function  $F$  of  $\theta$  can be defined on  $[-\pi, \pi]$  by the equation

$F(\theta_1) = P(-\pi < \theta \leq \theta_1)$ . This function  $F$  is usually extended to the whole real line by the equation

$F(\theta_1 + 2\pi) - F(\theta_1) = 1, -\infty < \theta_1 < \infty$ . The function  $F$  defined this way is called the distribution function (d.f.) of  $\theta$  on the circle.

Given two points  $\theta_1$  and  $\theta_2$  on  $S^1$ , we denote by  $\text{arc}(\theta_1, \theta_2)$  the set of points from  $\theta_1$  to  $\theta_2$  in the counter-clockwise direction with  $\theta_1$  excluded and  $\theta_2$  included. It follows that

$P(\theta \in \text{arc}(\theta_1, \theta_2)) = F(\theta_2) - F(\theta_1)$ . There is a natural projection from  $R^1$  to  $S^1$  defined by  $x \mapsto \theta = x \bmod 2\pi$ . Let  $\theta_1 = x_1 \bmod 2\pi$  and  $\theta_2 = x_2 \bmod 2\pi$ . It can be shown that

$P(\theta \in \text{arc}(\theta_1, \theta_2)) = (F(x_2) - F(x_1)) \bmod 1$ . We note that the d.f.  $F$  is a right continuous function, but in contrast with d.f.'s on the real line,

$$\lim_{\theta \rightarrow \infty} F(\theta) = \infty, \quad \lim_{\theta \rightarrow -\infty} F(\theta) = -\infty.$$

If the d.f.  $F$  is absolutely continuous, it has a probability density function (p.d.f.)  $f$  such that

$$\int_{\theta_1}^{\theta_2} f(\theta) d\theta = F(\theta_2) - F(\theta_1). \quad \text{A given function } f \text{ is the p.d.f.}$$

of an absolutely continuous distribution if and only if (i)  $f(x) \geq 0$ ,  $x \in R^1$ , (ii)  $f(x + 2\pi) = f(x)$ , (iii)  $\int_{-\pi}^{\pi} f(x) dx = 1$ .



## II. 2. The Characteristic Function.

Another representation of  $S^1$  is as the set of complex numbers of unit length. Any such number can be uniquely written as  $e^{i\theta}$ ,  $\theta \in (-\pi, \pi]$ . If  $\theta$  is a random variable taking values on  $(-\pi, \pi]$ , then  $z = \exp i\theta$  is a random variable taking values on the unit circle in the complex plane, which will also be denoted  $S^1$ .

The characteristic function of  $\theta$  (or  $z$ ) is the function  $\psi$  defined to integers,  $t=0, \pm 1, \pm 2, \dots$ , by  $\psi(t) = E \exp(it\theta) = \int_{-\pi}^{\pi} \exp(it\theta) dF(\theta)$  where  $F(\theta)$  is the circular d.f. of  $\theta$ . Obviously  $\psi(0) = 1, \psi(0) = 1, \psi(-p), |\psi(p)| \leq 1$ . The expectations,  $\alpha(t) = E(\cos t\theta) = \operatorname{Re} \psi(t)$  and  $\beta(t) = E(\sin t\theta) = \operatorname{Im} \psi(t)$ , are called the  $t$ -th order sine and cosine moments respectively. If  $\sum_{t=1}^{\infty} (\alpha^2(t) + \beta^2(t))$  is convergent, the random variable  $\theta$  has a density which is defined almost everywhere by

$$f(\theta) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \psi(t) \exp(-it\theta)$$

The joint c.f. of two circular random variables  $\theta_1$  and  $\theta_2$  is defined by  $\psi(t,s) = E \exp i(t\theta_1 + s\theta_2)$  where  $t$  and  $s$  are integers. Let the c.f.'s of  $\theta_1$  and  $\theta_2$  be  $\psi_1(t)$  and  $\psi_2(s)$ . Then  $\theta_1$  and  $\theta_2$  are independent if and only if  $\psi(t,s) = \psi_1(t) \psi_2(s)$ . Furthermore, if the p.d.f.'s of  $\theta_1$  and  $\theta_2$  exist, then their convolution is the p.d.f. of  $\theta_1 \oplus \theta_2$ , i.e.

$$(2\pi)^{-1} \sum \psi_1(t) \psi_2(t) \exp(-it\theta).$$

### II. 3. Error Criteria and Optimal Estimates.

The standard distance function on the circle, the distance  $\rho$  between two points on the circle, is the arc length of the short path joining them. If we restrict  $\theta_1$  and  $\theta_2$  to take values in the range  $(-\pi, \pi]$ , we have

$$\rho(\theta_1, \theta_2) = \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|).$$

The class of error criteria we wish to consider is the class of symmetric, nondecreasing cost functions--i.e. functions  $\phi: S^1 \rightarrow \mathbb{R}$  which satisfy

$$\begin{aligned} 0 \leq \phi(\theta) &= \phi(-\theta) \\ 0 \leq \rho(\theta_1, 0) \leq \rho(\theta_2, 0) &\Rightarrow \phi(\theta_1) \leq \phi(\theta_2). \end{aligned} \quad (1)$$

Some examples of cost criteria satisfying (1) are  $\rho(\theta) \triangleq \rho(\theta, 0)$ ,  $(1 - \cos\theta)$ ,  $\rho(\theta)^2$ ,  $(1 - \cos\theta)^2$ . We also wish to consider the special class of unimodal, mode-symmetric probability density functions--i.e., density functions of the form  $p: S^1 \rightarrow [0, \infty)$  with a unique maximum at  $\eta$ , such that

$$p(\eta + \theta) = p(\eta - \theta) \quad \forall \theta.$$

As the following theorem demonstrates, under these conditions the mode of the density is the optimal estimate.

Theorem 1: Given an error function  $\phi$  that satisfies (1) and a unimodal, mode-symmetric probability density function  $p$ , then the

estimation error is minimized at the mode, i.e.,

$$E(\phi(\theta-\eta)) \leq E(\phi(\theta-a)) \quad \forall a$$

where  $p$  has its maximum at  $\eta$ .

Proof: The theorem follows immediately from results on similarly ordered functions and the rearrangement inequalities. The basic result for real valued functions defined on  $R^1$  is contained in [18] (thm. 378) and [19, p. 183]. The result for  $S^1$  is obtained by making only minor changes in these proofs.

We remark that from the symmetry of the problem,  $\phi$  has its global maximum at  $\pi$  and  $p$  has its global minimum at  $\eta + \pi$ . Thus

$$E(\phi(\theta-\eta+\pi)) \geq E(\phi(\theta-\alpha)) \quad \forall \alpha. \quad \blacksquare$$

It should be noted that Theorem 1 is the  $S^1$  analog of a result of [20], [21]. Note that the same result is true if no probability density exists but the probability measure is unimodal at, and symmetric about, some point  $\eta$ , i.e., the d.f.  $F$  is convex for  $(-\pi, 0]$  and if  $F(\theta) = 1 - F(-\theta)$  at each continuity point of  $F$ .

Let us now restrict our attention to the error function,  $\phi(\theta) = 1 - \cos \theta$ . This function was used widely in statistics [4] and was used in [13] to design a phase-tracking system. It is especially interesting, because locally it is a quadratic function, i.e.  $1 - \cos \theta \approx 1/2 \theta^2$  for  $\theta \ll 1$ . Let  $\hat{\theta}$  denote the optimal estimate of the random variable  $\theta$  on  $S^1$  with respect to the error criterion



$E(1-\cos(\theta-\hat{\theta}))$ . As

$$\eta(\theta, \hat{\theta}) = E(1-\cos(\theta-\hat{\theta})) = 1 - [E \cos \theta, E \sin \theta] [\cos \hat{\theta}, \sin \hat{\theta}]^T,$$

the optimal estimate  $\hat{\theta}$  is determined by

$$\cos \hat{\theta} = \frac{1}{\rho} E \cos \theta,$$

$$\sin \hat{\theta} = \frac{1}{\rho} E \sin \theta,$$

with

$$\rho = [(E \cos \theta)^2 + (E \sin \theta)^2]^{\frac{1}{2}}.$$

We note that the complex number  $\psi(1)$  defined in Subsection II.2 can be expressed as  $\rho \exp i\hat{\theta}$ . This number is called the resultant of  $\theta$ . In analogy to the linear space case, the optimal estimate  $\hat{\theta}$  is called the circular mean of  $\theta$ , and the estimation error  $\eta = 1-\rho$  is called the circular variance.

#### II. 4. Folded Normal Densities.

Given a random variable  $x$  on  $R^1$  with d.f.  $F_x$ , the random variable  $\theta = x \bmod 2\pi$  on the circle has the d.f.  $F$  defined by

$$F(\theta) = \sum_{k=-\infty}^{\infty} (F_x(\theta+2\pi k) - F_x(2\pi k - \pi)), \quad \theta \in (-\pi, \pi].$$

This can be viewed as obtained from wrapping  $F_x$  around the circumference of the unit circle. If  $x$  has a p.d.f.  $p_x(x)$ , the corresponding p.d.f. of  $\theta$  is

$$p(\theta) = \sum_{k=-\infty}^{\infty} p_x(\theta + 2k\pi).$$

Corresponding to a normal density  $p_x$ , the folded normal density

$$F(\theta; \eta, \gamma) = \frac{1}{\sqrt{2\pi\gamma}} \sum_{k=-\infty}^{\infty} \exp \left[ -\frac{1}{2\gamma} (\theta - \eta + 2k\pi)^2 \right] \quad (2a)$$

plays a central role in the continuous-time estimation problem considered in Section IV. The Fourier series representation of the folded normal density is

$$F(\theta; \eta, \gamma) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \exp \left( -\frac{k^2 \gamma}{2} \right) \cos k(\theta - \eta). \quad (2b)$$

From this representation it is easy to see that the convolution of two folded normal densities,  $F(\theta; \eta_1, \gamma_1)$  and  $F(\theta; \eta_2, \gamma_2)$ , is the folded normal density  $F(\theta; \eta_1 \oplus \eta_2, \gamma_1 + \gamma_2)$ . More important properties of the folded normal density will be studied in the following [6].

**Theorem 2:** The folded normal density, (2a) and (2b), is unimodal with mode at  $\theta = \eta$  and is symmetric about  $\eta$ .

**Proof:** Since  $\cos \leq 1$ , the second form of  $F$  in (2b) yields

$$F(\theta; \eta, \gamma) \leq \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-n^2 \gamma/2} = F(\eta; \eta, \gamma).$$

Thus  $F$  has its global maximum at  $\theta = \eta$ .

Since  $F(\theta, \eta, \gamma) = F(\theta - \eta; 0, \gamma)$ , we need only show that  $F(\theta; 0, \gamma)$  is symmetric about 0 and monotone decreasing as  $\rho(\theta, 0)$  increases. Symmetry is obvious ( $\cos n\theta = \cos n(-\theta)$ ), and monotonicity will follow if we can show

$$\frac{\partial F}{\partial \theta}(\theta; 0, \gamma) < 0 \quad \theta \in (0, \pi) \quad (3a)$$



$$\frac{\partial F}{\partial \theta}(\theta; 0, \gamma) > 0 \quad \theta \in (-\pi, 0) \quad (3b)$$

We now remark that the properties of  $F(\theta; 0, \gamma)$  have been studied extensively, since it is a theta function. See [22] and [23] for discussions of some properties of theta functions. Using the notation of [22, pp. 2, 42], we have

$$F(\theta; 0, \gamma) = \frac{1}{2\pi} \theta_4 \left( \frac{\theta + \pi}{2}, \frac{i\gamma}{2\pi} \right) = k \prod_{n=1}^{\infty} (1 + 2q^{2n-1} \cos \theta + q^{4n-2}), \quad (4)$$

where  $q = e^{-\gamma/2}$  and

$$k = \frac{1}{2\pi} \prod_{n=1}^{\infty} (1 - q^{2n}).$$

Using the fact that  $F > 0$  and the form of  $F$  given by (4) we have

$$F^{-1}(\theta; 0, \gamma) \frac{\partial F}{\partial \theta}(\theta; 0, \gamma) = - \left[ \sum_{n=1}^{\infty} \frac{2q^{2n-1}}{(1 + 2q^{2n-1} \cos \theta + q^{4n-2})} \right] \sin \theta. \quad (5)$$

It is easily seen that the term in square brackets on the right hand side of (5) is positive for all values of  $\theta$  and thus (3) is correct.

Some work along these lines has been done in [53]. See [23] for discussions of other relevant properties of theta functions, hypergeometric functions, Legendre polynomials, and Tchebycheff polynomials.

Note that the symmetry requirements of Theorem 1 are necessary. For instance, if  $\phi$  is not symmetric, the mode of the density need not be the optimal estimate even if all the other assumptions of Theorem 1 do hold. As an example, consider the function  $\phi : S^1 \rightarrow \mathbb{R}$

$$\phi(\theta) = \begin{cases} \theta & 0 \leq \theta \leq \pi \\ \frac{\theta^2}{\pi} & -\pi \leq \theta \leq 0. \end{cases}$$

Suppose our distribution is the folded normal centered at 0. Then it can be shown that the mode, 0, is not the optimal estimate.

**Theorem 3.** Let  $\phi$  satisfy the second requirement of (1) and let  $p(\theta) = F(\theta; \eta, \gamma)$ . Then  $E(\phi(\theta - \eta))$  is an increasing function of the variance,  $\gamma$  -- that is

$$\frac{d}{d\gamma} E(\phi(\theta - \eta)) \geq 0. \quad (6)$$

Proof: Writing

$$\phi(\theta) = d_0 + \sum_{n=1}^{\infty} c_n \sin n\theta + d_n \cos n\theta$$

and using the results on Fourier series analysis,

$$E(\phi(\theta - \eta)) = d_0 + \sum_{n=1}^{\infty} d_n e^{-n^2 \gamma / 2}, \quad (7)$$

but we get the same error if we compute  $E(\psi(\theta - \eta))$ , where  $\psi$  is the symmetrized function

$$\psi(\theta) = [\phi(\theta) + \phi(-\theta)]/2.$$

which also satisfies (1). Thus, it is enough to prove the theorem for  $\phi$  satisfying (1). In this case  $\eta$  is the optimal estimate and

$$\begin{aligned} E(\phi(\theta - \eta)) &= \int \phi(\theta - \eta) F(\theta, \eta, \gamma) d\theta \\ &= \int_{-\pi}^{\pi} \phi(\theta) F(\theta; 0, \gamma) d\theta \\ &= 2 \int_0^{\pi} \phi(\theta) F(\theta; 0, \gamma) d\theta. \end{aligned}$$

Then, (6) will hold if

$$\int_0^\pi \phi(\theta) \frac{\partial}{\partial \gamma} F(\theta; 0, \gamma) d\theta \geq 0$$

Suppose we can show that there exists  $\theta_0 \in [0, \pi]$  such that

$$\frac{\partial}{\partial \gamma} F(\theta; 0, \gamma) < 0 \quad \theta \in (0, \theta_0)$$

$$\frac{\partial}{\partial \gamma} F(\theta_0; 0, \gamma) = 0$$

$$\frac{\partial}{\partial \gamma} F(\theta; 0, \gamma) > 0 \quad \theta \in (\theta_0, \pi).$$

Then, since

$$\phi(\theta) \leq \phi(\theta_0) \quad \theta \in [0, \theta_0]$$

$$\phi(\theta) \geq \phi(\theta_0) \quad \theta \in [\theta_0, \pi],$$

we have

$$\int_0^\pi \phi(\theta) \frac{\partial}{\partial \gamma} F(\theta; 0, \gamma) d\theta \geq \phi(\theta_0) \frac{d}{d\gamma} \int_0^\pi F(\theta; 0, \gamma) d\theta$$

$$= \phi(\theta_0) \frac{d}{d\gamma} (1/2) = 0,$$

and we get a strict inequality if  $\phi$  is not a constant.

Now it is easy to see that

$$\frac{\partial}{\partial \gamma} F(\theta; 0, \gamma) = 1/2 \frac{\partial^2}{\partial \theta^2} F(\theta; 0, \gamma)$$

and the theorem will be proved once we prove the following lemma, which yields more information about the shape of the folded normal density.

**Lemma 1:** For an arbitrary but fixed value of  $\gamma > 0$ , there exists

$\theta_0 \in [0, \pi]$  such that



$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} F(\theta; 0, \gamma) &< 0 & \theta \in [0, \theta_0) \\ \frac{\partial^2}{\partial \theta^2} F(\theta_0; 0, \gamma) &= 0 \\ \frac{\partial^2}{\partial \theta^2} F(\theta; 0, \gamma) &> 0 & \theta \in (\theta_0, \pi] \end{aligned}$$

That is,  $F$  has a unique inflection point (at  $\theta_0$ ) on  $[0, \pi]$ .

Proof: We use the form of  $F(\theta; 0, \gamma)$  given in equation (4). We compute

$$F^{-1} \frac{\partial^2}{\partial \theta^2} F = -A \cos \theta + B \sin^2 \theta$$

$$A = \sum_{n=1}^{\infty} \frac{2q^{2n-1}}{(1+2q^{2n-1} \cos \theta + q^{4n-2})}$$

$$B = \sum_{n \neq m} \frac{4q^{2(n+m-1)}}{(1+2q^{2n-1} \cos \theta + q^{4n-2})(1+2q^{2m-2} \cos \theta + q^{4m-2})}$$

and then a simple computation yields

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} F(0; 0, \gamma) &< 0 \\ \frac{\partial^2}{\partial \theta^2} F(\theta; 0, \gamma) &> 0 & \forall \theta \in [\frac{\pi}{2}, \pi] \end{aligned}$$

and

$$\frac{\partial}{\partial \theta} \left( F^{-1} \frac{\partial^2 F}{\partial \theta^2} \right) (\theta; 0, \gamma) > 0 \quad \forall \theta \in (0, \frac{\pi}{2}).$$

These inequalities imply that, there is a unique  $\theta_0 \in (0, \frac{\pi}{2})$  such that

$$\frac{\partial^2}{\partial \theta^2} F(\theta_0; 0, \gamma) = 0 \text{ and that } F(\theta; 0, \gamma) > 0 \text{ for } \theta > \theta_0.$$

$$\frac{\frac{\partial^2}{\partial \theta^2} F(\theta_1; 0, \gamma)}{F(\theta_1; 0, \gamma)} > \frac{\frac{\partial^2}{\partial \theta^2} F(\theta_0; 0, \gamma)}{F(\theta_0; 0, \gamma)} = 0$$

$$\text{or } \frac{\partial^2}{\partial \theta^2} F(\theta_1; 0, \gamma) > 0$$

and the lemma and the theorem are proved.

Note that by symmetry we have that  $F$  has a unique inflection point at  $-\theta_0$  on the interval  $[-\pi, 0]$ .

Theorem 3 tells us that the intuitive notion that we "have more accurate information" for smaller values of  $\gamma$  can be made precise. Also, this theorem implies another result, which is the  $S^1$  analog of a problem treated in [24]. The problem treated in [24] is that of finding the optimal linear filter minimizing an asymmetric error criterion on  $R^1$  that decreases on  $(-\infty, 0]$  and increases on  $[0, \infty)$ . The result is that the optimal linear filter is the minimum variance filter, and the proof essentially consists of showing that the expected error cost is an increasing function of the variance. Theorem 3 clearly implies an  $S^1$  analog of this result.

Some examples of cost criteria satisfying (1) and the associated optimal costs when the density is folded-normal are given in the following theorem, of which the proof is simple and is therefore omitted.

**Theorem 4.** Let  $p(\theta) = F(\theta; \eta, \gamma)$ . Then

$$(i) \quad E(1 - \cos(\theta - \eta)) = 1 - \exp\left(-\frac{\gamma}{2}\right)$$

$$(ii) \quad E(1 - \cos(\theta - \eta))^2 = \frac{3}{2} - 2 \exp\left(-\frac{\gamma}{2}\right) + \frac{1}{2} \exp(-2\gamma)$$



$$(iii) \quad E(\rho(\theta-\eta)) = \frac{1}{2} - \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} \exp(-\frac{1}{2}(2k+1)^2 \gamma)$$

$$(iv) \quad E(\rho^2(\theta-\eta)) = \frac{2\pi^2}{3} - 4 \left[ \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2} \exp(-\frac{k^2 \gamma}{2}) \right]$$

## II. 5. Circular Normal Densities.

The multistage estimation for discrete-time systems on  $S^1$  involves two operations alternately that are convolution and conditioning (i.e. taking conditional distribution). While the class of folded normal densities is closed under convolution, unfortunately it is not closed under conditioning. The difficulty involved in using folded normal densities for discrete-time estimation was discussed in [6] and [25].

The difficulty is partially resolved [5] if another class of "normal" densities on  $S^1$ , is used. These densities are called circular normal densities and have the form

$$G(\theta; \eta, \gamma) = \frac{1}{2\pi I_0(\gamma)} \exp \gamma \cos (\theta - \eta),$$

where  $I_0(\gamma)$  is the modified Bessel function of the first kind and order zero, i.e.,

$$I_0(\gamma) = \sum_{k=0}^{\infty} (\gamma/2)^{2k} / (k!)^2.$$

The circular normal density was first introduced by Langevin [26] in 1905 and by Von Mises [27] in 1918 in the context of statistical mechanics. In

contrast to the folded normal densities, the class of circular normal densities is closed under conditioning rather than under convolution [5]. More will be said about this in the next section after the circular normal density is generalized to the exponential Fourier density.

The class of normal densities on an Euclidean space has the closure properties under both convolution and conditioning, which accounts for the success of the Kalman-Bucy filtering for the discrete-time systems. Now the folded and the circular normal densities divide these two properties between them. Which one then, is more "normal" than the other? We recall that the linear normal density has two characterizations -- the maximum likelihood characterization and the maximum entropy characterization. It was observed by Von Mises [27] and Mardia [4] respectively that the circular normal density has both characterizations on the circle. However, the Brownian motion on the circle, induced by that on the real line through "rolling without slipping" (See Section IV), and a variant form of the central limit theorem on the circle (See [4]) both lead to the folded normal density. Further, the independence of  $p(\theta_1)$  and  $p(\theta_2)$  and  $p(\theta_1) - p(\theta_2)$ , where  $p$  is an arbitrary function and  $\theta_1$  and  $\theta_2$  are independent, also leads to the folded normal density (See [4]). Therefore, there may be no answer to the above question. Before we start the next section, let us have a few words about the shape of the circular normal density.

The circular normal density  $G(\theta; \eta, \gamma)$  is obviously unimodal and symmetric about the mode  $\eta$ . The ratio of the density at the mode to that at the antimode  $\eta + \pi$  is given by  $\exp 2\gamma$  so that the larger the

value of  $\gamma$ , the greater is the clustering around the mode. It can be shown by straightforward calculation that the function  $G(\theta; 0, \gamma)$  has two inflection points, at  $\pm \arccos [-\gamma/2 + (1 + \gamma^2/4)^{1/2}]$ .

### III. DISCRETE-TIME ESTIMATION ON THE CIRCLE.

Estimation for discrete-time systems on the circle was studied in [6] and [25], using both folded normal densities and Fourier series representations of probability densities. The optimal estimation equations obtained therein are infinite-dimensional and cumbersome. Although some numerical simulation has been done on the suboptimal equations obtained through truncating the higher order terms, it is not clear whether these equations have satisfactory performance in general.

As a matter of fact, the "dimension" of the optimal estimation equations derived from using the folded normal densities increases very rapidly in time. When Fourier series are used to represent probability densities, the application of Bayes' rule, which involves the multiplication of two a priori densities, has the effect of spreading the dominant Fourier coefficients into the higher order terms. Obviously, this dilemma becomes compounded in a multistage estimation problem when a sequence of multiplications of Fourier series takes place.

In this section, we will present an alternative approach. The approach is based on a new class of probability density functions which have the form

$$\exp \left[ \sum_{k=0}^n (a_k \cos kx + b_k \sin kx) \right].$$

Such a density will be called an exponential density of order  $n$ , to be denoted by  $EFD(n)$ . We note that the circular normal density introduced



in the previous section is exactly the EFD(1).

### III. 1. Exponential Fourier Densities on the Circle.

There are two reasons for using the exponential Fourier densities. It is obvious that the multiplication of two EFD(n)'s does not raise the order of the densities. Thus the class of n-th order exponential Fourier densities is closed under the operation of taking conditional distributions.

Another reason for using EFD's is that any continuous or bounded variation function can be approximated by an EFD as closely as desired with respect to the square integral norm. This property enables us to use an EFD(n) as a mathematical model of any probability distribution on the circle. Both this and the aforementioned closure properties can be generalized to compact Lie groups and some homogeneous spaces, as will be seen in Section V.

Before we illustrate how the EFD's are used to deduce finite-dimensional, closed-form, and recursive equations to update the conditional densities of the signal given the observation, we will now state the approximation property in the following theorem of which a general version for compact Lie groups will be proven in Section V.

Theorem. Let  $p$  be a continuous probability density on  $S^1$ . For any given positive number  $\epsilon$ , there exists an exponential Fourier density,  $p_n(x) = \exp \sum_{k=0}^n (a_k \cos kx + b_k \sin kx)$ , such that

$$\int_{-\pi}^{\pi} (p(x) - p_n(x))^2 dx \leq \epsilon.$$

### III. 2. A Basic System on $S^1$ .

Assume that the signal and the measurement processes are governed by the equations

$$s_{n+1} = s_n \oplus w_n$$

$$m_n = s_n \oplus v_n$$

where  $\{w_n\}$  is a given deterministic process on  $S^1$ , and  $\{v_n\}$  is a white random process on  $S^1$ . The probability densities of  $s_1$  and  $v_n$  are assumed to be the following independent exponential Fourier densities:

$$p(s_o) = \exp \sum_{k=0}^N (a_{ok} \cos ks_1 + b_{ok} \sin ks_o)$$

$$p(v_n) = \exp \sum_{k=0}^N (\alpha_{nk} \cos kv_n + \beta_{nk} \sin kv_n).$$

By Bayes' rule,

$$p(s_{n+1} | m^{n+1}) = c_{n+1} p(m_{n+1} | s_{n+1}) p(s_{n+1} | m^n) \quad (8)$$

with  $c_{n+1} = 1/p(m_{n+1} | m^n)$  = a normalizing constant. It can be easily shown that the conditional densities on the right can be written as the following exponential Fourier densities:

$$p(s_n | m^n) = \exp \sum_{k=0}^N (a_{nk} \cos ks_n + b_{nk} \sin ks_n) \quad (9)$$

$$p(s_{n+1} | m^n) = \exp \sum_{k=0}^N (a_{nk} \cos k(s_{n+1} - w_n) + b_{nk} \sin k(s_{n+1} - w_n)) \quad (10)$$

$$p(m_{n+1} | s_{n+1}) = \exp \sum_{k=0}^N (\alpha_{n+1,k} \cos k(m_{n+1} - s_{n+1}) + \beta_{n+1,k} \sin k(m_{n+1} - s_{n+1})),$$

where  $a_{nk}$  and  $b_{nk}$  are to be determined. Substituting these two equations into (3),

$$\begin{aligned}
 p(s_{n+1} | m^{n+1}) = c_{n+1} \exp \sum_{k=0}^N [ & (a_{nk} \cos kw_n \\
 & - b_{nk} \sin kw_n + \alpha_{n+1,k} \cos km_{n+1} \\
 & + \beta_{n+1,k} \sin km_{n+1}) \cos ks_{n+1} + (a_{nk} \sin kw_n \\
 & + b_{nk} \cos kw_n + \alpha_{n+1,k} \sin km_{n+1} \\
 & - \beta_{n+1,k} \cos km_{n+1}) \sin ks_{n+1} ].
 \end{aligned}$$

Thus, we obtain the following recursive formulas for  $a_{nk}$  and  $b_{nk}$  which, in turn, give us the desired conditional densities  $p(s_n | m^n)$ :

$$\begin{aligned}
 a_{n+1,k} &= a_{nk} \cos kw_n - b_{nk} \sin kw_n \\
 &+ \alpha_{n+1,k} \cos km_{n+1} + \beta_{n+1,k} \sin km_{n+1} \\
 b_{n+1,k} &= a_{nk} \sin kw_n + b_{nk} \cos kw_n \\
 &+ \alpha_{n+1,k} \sin km_{n+1} - \beta_{n+1,k} \cos km_{n+1} \\
 p(s_{n+1} | m^{n+1}) &= \exp \sum_{k=0}^N (a_{n+1,k} \cos ks_{n+1} + b_{n+1,k} \sin ks_{n+1}),
 \end{aligned}$$

for  $k = 1, 2, \dots$ , and where  $a_{n+1,0}$  is a normalizing constant.



### III. 3. A Phase-Shift-Keyed System.

Consider the signal and the measurement processes governed by the equations

$$s_{n+1} = s_n \oplus w_n$$

$$m_n = \cos(\omega t_n + s_n) + v_n$$

where  $\{w_n\}$  is a given deterministic process on  $S^1$  and  $\{v_n\}$  is a white Gaussian sequence with zero mean and variances  $\sigma_n^2$ . The probability density of  $s_1$  is assumed to be the exponential Fourier density

$$p(s_1) = \exp \sum_{k=0}^N (a_{1k} \cos ks_1 + b_{1k} \sin ks_1).$$

We note that the measurement process  $\{m_n\}$  can be viewed as a sampled sinusoidal wave modulated by a random phase process  $\{s_n\}$  and corrupted by additive white Gaussian noise  $\{v_n\}$ . The special case of this model where  $p(s_1)$  is a first-order exponential Fourier density has been solved in [13]. Here again, by Bayes' rule and straightforward calculation, we have (8-10). As  $v_{n+1}$  is a Gaussian random variable, it follows that:

$$p(m_{n+1} | s_{n+1}) = \frac{1}{\sqrt{2\pi} \sigma_{n+1}} \exp \left[ - \frac{(m_{n+1} - \cos(\omega t_{n+1} + s_{n+1}))^2}{2\sigma_{n+1}^2} \right] \quad (11)$$

Substituting (11) and (10) into (8) yields

$$\begin{aligned}
 p(s_{n+1} | m^{n+1}) &= \frac{c_{n+1}}{\sqrt{2\pi} \sigma_{n+1}} \exp \sum_{k=0}^N [ a_{nk} \cos k(s_{n+1} - w_n) \\
 &\quad + b_{nk} \sin k(s_{n+1} - w_n) \\
 &\quad - \frac{1}{2\sigma_{n+1}^2} (m_{n+1}^2 - 2m_{n+1} \cos(\omega t_{n+1} + s_{n+1}) \\
 &\quad + \cos^2(\omega t_{n+1} + s_{n+1})) ] \\
 &= \frac{c_{n+1}}{\sqrt{2\pi} \sigma_{n+1}} \exp \{ \sum_{k=0}^N [(a_{nk} \cos kw_n - b_{nk} \sin kw_n) \cos ks_{n+1} \\
 &\quad + (a_{nk} \sin kw_n + b_{nk} \cos kw_n) \sin ks_{n+1} \\
 &\quad + \frac{m_{n+1}}{\sigma_{n+1}^2} \cos \omega t_{n+1} \cos s_{n+1} \\
 &\quad - \frac{m_{n+1}}{\sigma_{n+1}^2} \sin \omega t_{n+1} \sin s_{n+1} - \frac{1}{4\sigma_{n+1}^2} \cos 2\omega t_{n+1} \\
 &\quad \cdot \cos 2s_{n+1} + \frac{1}{4\sigma_{n+1}^2} \sin 2\omega t_{n+1} \sin 2s_{n+1} - \frac{m_{n+1}^2}{2\sigma_{n+1}^2} - \frac{1}{2} \} .
 \end{aligned}$$

Thus we obtain the following recursive formulas for  $a_{nk}$  and  $b_{nk}$  which, in turn, give us the desired conditional densities  $p(s_n | m^n)$ :

$$a_{n+1,1} = a_{n1} \cos w_n - b_{n1} \sin w_n + \frac{m_{n+1}}{\sigma_{n+1}^2} \cos \omega t_{n+1},$$

$$b_{n+1,1} = a_{n1} \sin w_n + b_{n1} \cos w_n - \frac{m_{n+1}}{\sigma_{n+1}^2} \sin \omega t_{n+1},$$

$$a_{n+1,2} = a_{n2} \cos 2w_{n2} - b_{n2} \sin 2w_n - \frac{1}{4\sigma_{n+1}^2} \cos 2\omega t_{n+1},$$

$$b_{n+1,2} = a_{n2} \sin 2w_n + b_{n2} \cos 2w_n + \frac{1}{4\sigma_{n+1}^2} \sin 2\omega t_{n+1};$$

and, for  $k = 3, 4, \dots$ , recursively

$$a_{n+1,k} = a_{nk} \cos kw_n - b_{nk} \sin kw_n,$$

$$b_{n+1,k} = a_{nk} \sin kw_n + b_{nk} \cos kw_n,$$

$$p(s_{n+1} | m^{n+1}) = \exp \sum_{k=0}^N [a_{n+1,k} \cos ks_{n+1} + b_{n+1,k} \sin ks_{n+1}],$$

where  $a_{n+1,0}$  is a normalizing constant.

#### III. 4. Periodic Measurements in Additive White Gaussian Noise

Consider the signal and measurement processes

$$s_{n+1} = s_n \oplus w_n$$

$$m_n = h(s_n) + v_n$$

where  $\{w_n\}$  and  $\{v_n\}$  are as in the previous section and where  $h$  is a periodic function with a period of  $2\pi$ .

The periodicity of  $h$  allows us to approximate it by a finite Fourier series, as closely as we wish in the space of square-integrable functions. In other words, for any  $\epsilon > 0$ , there exist  $\{f_k\}, \{g_k\}$ , and a positive integer  $M$  such that



$$||h - h_M|| < \varepsilon$$

where

$$h_M(s) = \sum_{k=0}^M (f_k \cos ks + g_k \sin ks). \quad (12)$$

Without loss of generality, we may assume that  $N \geq 2M$  in (12), for otherwise we can set  $a_{1k} = b_{1k} = 0$ , for  $N \leq k \leq 2M$ , and write  $p(s_1) = \exp \sum_{k=0}^{2M} (a_{1k} \cos ks_1 + b_{1k} \sin ks_1)$ . We can also assume that  $f_0 = 0$ , for otherwise  $f_0$  can be incorporated into  $m_n$ . Assume that  $p(s_n | m^n) = \exp \sum_{k=0}^N (a_{nk} \cos ks_n + b_{nk} \sin ks_n)$ . By Bayes' rule and straightforward calculation, we obtain

$$\begin{aligned} p(s_{n+1} | m^{n+1}) &= c_{n+1} p(m_{n+1} | s_{n+1}) p(s_{n+1} | m^n) \\ &= \frac{c_{n+1}}{\sqrt{2\pi} \sigma_{n+1}} \exp \left[ -\frac{1}{2\sigma_{n+1}^2} \left( m_{n+1} - \sum_{k=0}^M (f_k \cos ks_{n+1} + g_k \sin ks_{n+1}) \right)^2 \right. \\ &\quad \left. + \sum_{k=0}^N (a_{nk} \cos k(s_{n+1} - w_n) + b_{nk} \sin k(s_{n+1} - w_n)) \right]. \end{aligned} \quad (13)$$

We note that the function in the above bracket can be written as a finite Fourier series of order  $N$  in the variable  $s_{n+1}$ . This shows by induction that for all  $n = 1, 2, \dots$ ,  $p(s_n | m^n)$  is an exponential Fourier density of order  $N$ ; the recursive formulas for  $a_{nk}$  and  $b_{nk}$  can be straightforwardly obtained from (13). However, the formulas are tedious and will not be displayed here.

#### IV. CONTINUOUS-TIME ESTIMATION ON THE CIRCLE.

A signal process and an observation process, taking values on  $S^1$ , will be formulated in terms of bilinear Ito matrix differential equations. The conditional probability distribution of the signal, given observations over a certain period of time, will be evaluated. Recursive computational schemes for optimal estimation (filtering, smoothing, and prediction), with respect to the error criteria defined in Subsection II.3, will be derived. In fact it will be shown that optimal estimates on  $S^1$  can be obtained recursively by the use of an ordinary vector space estimator together with a nonlinear preprocessor and a nonlinear postprocessor. Multichannel estimation on abelian Lie groups will be examined. Examples illustrating the optimal estimation procedure are given at the end of this section.

##### IV. 1. Signal Processes and Observation Processes

Consider the situation of a unit circle in  $R^2$  with a line tangent to it.

We allow the line to perform a one-dimensional continuous translation (along itself); fix the center of the circle and require that there be no slipping at the point of tangency. The line then induces a rotation of the circle and if the line moves a distance  $x$  the circle rotates  $x$  radians and so is  $x \bmod 2\pi = \theta$  radians away from its initial orientation.

This method, called "rolling without slipping", will now be used to construct a continuous signal process on  $S^1$  and to formulate the mathematical model of a sensor (i.e., an observation process) to be used in this report.

We will adopt the following notation

$(\Omega, A, P)$  = a probability space

$s$  = a positive real number

$C_1^s$  = the family of real-valued continuous functions,  $a$ , on  $[0, s]$  such that  $a(0) = 0$

$B_1^s$  = the Borel  $\sigma$ -field of  $C_1^s$

$C_2^s$  = the family of  $2 \times 2$  orthogonal-matrix-valued continuous functions,  $A$ , on  $[0, s]$  such that  $A(0)$  is the identity matrix  $I$ ,

$B_2^s$  = the Borel  $\sigma$ -field of  $C_2^s$ ,

$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ .

Lower case letters denote elements in  $C_1^s$  and upper case letters denote elements in  $C_2^s$ .

Let  $J: C_1^s \rightarrow C_2^s$  be defined by

$$(J(a))(t) = \exp(a(t)R) = \begin{bmatrix} \cos a(t) & \sin a(t) \\ -\sin a(t) & \cos a(t) \end{bmatrix} \quad (14)$$

for  $a \in C_1^s$  and  $t \in [0, s]$ . It is easily seen that  $J$  is  $B_1^s$ -measurable and bijective. A point on the unit circle,  $S^1$ , can be represented by either the angle  $\theta \in [-\pi, \pi)$  it makes with a fixed radial axis or the  $2 \times 2$  orthogonal matrix  $\exp(R\theta)$ . Therefore, in the first representation,  $C_2^s$  is



the family of piecewise continuous functions  $\theta(t)$ , such that at any point of discontinuity the right hand limit of  $\theta$  is  $+\pi$ , while the left-hand limit is  $-\pi$ .

Each continuous curve  $a(t)$  on  $R^1$  gives rise to one and only one piecewise continuous curve  $\theta(t)$  lying between  $\pi$  and  $-\pi$ , of which the continuous segments are obtained by translating the corresponding segments of  $a(t)$  by an integral multiple of  $2\pi$ . Conversely, each piecewise continuous curve in  $C_2^S$  gives rise to one and only one continuous curve taking values on  $R^1$  which is obtained simply by piecing the continuous segments together. This intuitive observation illustrates the bijective property of the operator  $J$ . Thus, a continuous random signal process on  $S^1$  which is described by an  $A$ -measurable function  $X: \Omega \rightarrow C_2^S$  corresponds to a continuous random signal process on  $R^1$  which is described by an  $A$ -measurable function  $x: \Omega \rightarrow C_1^S$  such that

$$X(t) = (J(x))(t), \quad t \in [0, s].$$

We now define a random process  $z: \Omega \rightarrow C_1^S$  by the K. Ito random differential equation,

$$dz(t) = m(x(t), t) dt + q^{1/2} dw(t), \quad z(0) = 0,$$

where  $m: R_1 \times R_1 \rightarrow R_1$  is Borel-measurable,  $q: R_1 \rightarrow R_1$  is positive and measurable and  $w$  is the standard Brownian motion on  $(\Omega, A, P)$ , independent of  $x$ . Let  $Z: \Omega \rightarrow C_2^S$  be defined by

$$Z(t) = (J(z))(t).$$

Applying the Ito differentiation rule, we obtain the following Ito matrix differential equation:

$$dZ(t) = Z(t) \begin{bmatrix} -\frac{q(t)}{2} & m(t) \\ -m(t) & -\frac{q(t)}{2} \end{bmatrix} dt + Z(t) \begin{bmatrix} 0 & dw(t) \\ -dw(t) & 0 \end{bmatrix} \quad (15)$$

$$Z(0) = I,$$

where  $m(t) \triangleq m(x(t), t)$  and the diagonal terms  $\frac{q(t)}{2}$  are the second order correction terms which keep  $Z$  on the circle. This equation is the mathematical model of the sensor to be used. We note that the input,  $x(t)$  to the sensor is not the dynamical state  $X(t)$  of the rotational signal process on the circle, but rather the angle the rotational process has swept.

The physical motivation for this sensor model comes from the fact that in observing a rotational process (for instance a gyroscope recording rotation about a fixed axis) our measurement contains information on the total rotation,  $x(t)$ , not just the orientation,  $X(t)$ . In some applications, such as the gyro problem mentioned above, we wish to extract knowledge of orientation from knowledge of rotation, so it is proper to regard  $X(t)$  as the signal process. However, in other applications, such as FM demodulation, our interest centers on the  $x$  process, and in these cases, we may regard  $x$  as the signal.

#### IV. 2. Conditional Probability Distributions.

In this subsection, we will derive equations for the conditional probability distribution of the signal process given observations over some time period. The approach of this section is measure-theoretic in nature, and the major results are summarized in the statements of Lemma 2 and of Theorem 5 and its two corollaries.

Let us denote  $\{z(\tau), \tau \in [0, t]\}$  and  $\{Z(\tau), \tau \in [0, t]\}$  by  $z^t$  and  $Z^t$ , respectively. We note that  $Z^t = J(z^t)$ . Since  $J$  is bijective from  $C_1^t$  to  $C_2^t$ , the  $\sigma$ -subfield of  $A$  generated by  $z^t$  is the same as that generated by  $Z^t$ . In other words, the information carried by  $z^t$  and  $Z^t$  is the same. That  $\sigma$ -subfield will be denoted by  $A_z^t$ . The  $\sigma$ -subfield of  $A$  which is generated by  $X_\lambda = X(\lambda)$  (the subscripts  $\lambda, s, t$  denote that the processes are evaluated at these times) will be denoted by  $A_{X_\lambda}$ .

Let  $P_{zx}$  be the conditional probability measure on  $(\Omega, A_x)$  given  $A_z^t$ , defined by  $P_{zx}(A, \omega_2) = P(A | A_z^t)(\omega_2)$ , for  $A \in A_x$ ,  $\omega_2 \in \Omega$ . Let  $P_{zx}$  be the conditional probability measure on  $(\Omega, A_z^t)$  given  $A_x$ , defined by

$P_{zx}(B, \omega_1) = P(B | A_x)(\omega_1)$ , for  $B \in A_z^t$ ,  $\omega_1 \in \Omega$ . The restrictions of  $P$  to  $A_z^t$  and  $A_x$  are denoted by  $P_z$  and  $P_x$ , respectively. Let  $\mu_z$  and  $\mu_w$  be the measures induced on  $(C_1^t, B_1^t)$  by  $z^t$  and  $w^t$ , respectively. Define the conditional measures  $\mu_{zx}$  on  $(C_1^t, B_1^t)$ , given  $X_\lambda$ , by  $\mu_{zx}(B, \omega_1) = P(z^{-1}(B) | A_x)(\omega_1)$ , for  $B \in B_1^t$ ,  $\omega_1 \in \Omega$ .

It is known [28] that  $\mu_{zx} \approx \mu_w \approx \mu_z$  where  $\approx$  denotes equivalence of measures, and that

$$\frac{d\mu_{zx}}{d\mu_w}(\xi^t, A_\lambda) = E_x[\theta^t | X_\lambda = A_\lambda]$$

$$\frac{d\mu_z}{d\mu_w}(\xi^t) = E_x[\theta^t]$$

Here  $E_x$  means taking the average over  $x$ . Further,

$$\theta = \exp \left( -1/2 \int_0^t \frac{m^2}{q}(\tau) d\tau + \int_0^t \frac{m}{q}(\tau) d\xi(\tau) \right)$$



where  $\int$  denotes an Ito integral. Hence

$$\frac{dP_{zx}}{dP_z}(\omega_2, \omega_1) = \frac{d\mu_{zx}}{d\mu_z}(z^t(\omega_2), X_\lambda(\omega_1)) = \frac{E_x(\theta^t | X_\lambda = X_\lambda(\omega_1))}{E_x(\theta^t)}$$

where

$$\theta^t = \exp \left( -1/2 \int_0^t \frac{m^2(\tau)}{q} d\tau + \int_0^t \frac{m}{q}(\tau) dz(\tau, \omega_1) \right). \quad (16)$$

We note that  $\frac{dP_{zx}}{dP_x}(\omega_2, \omega_1)$  is  $A_z \times A_x$ -measurable. Applying a general

Bayes rule from [29], we obtain

$$\frac{dP_{xz}}{dP_x}(\omega_1, \omega_2) = \frac{dP_{zx}}{dP_z}(\omega_2, \omega_1).$$

Let us denote the family of  $2 \times 2$  orthogonal matrices by  $M_0$  and the set of induced Borel sets by  $B_0$ . Let  $\mu_{xz}$  be the conditional measure on  $(M_0, B_0)$  given  $A_z$ , defined by  $\nu_{xz}(A, \omega_2) = P(X_\lambda^{-1}(A) | A_z^t)(\omega_2)$ , for  $A \in B_0, \omega_2 \in \Omega$ . Let  $\nu_x$  be the measure on  $(M_0, B_0)$  induced by  $X_\lambda$ . Then it is easily seen that

$$\frac{d\nu_{xz}}{d\nu_x}(X_\lambda(\omega_1), Z^t(\omega_2)) = \frac{dP_{xz}}{dP_x}(\omega_1, \omega_2) = \frac{E_x(\theta^t | X_\lambda = X_\lambda(\omega_1))}{E_x(\theta^t)}$$

where  $\theta^t$  is defined by (16). Summarizing what has been shown, we have the following lemma.

**Lemma 2:** Consider the observation process described by (15). The conditional probability measure  $\nu_{xz}$ , for the signal  $X_\lambda$  given the observation  $Z^t$  is then absolutely continuous with respect to  $\nu_x$ , the a priori measure for  $X_\lambda$ . Further, for  $Z^t \in C_2^t$  and  $X \in M_0$ , one has

$$\frac{dv_{xz}}{dv_x}(X, Z^t) = \frac{E_x(\theta^t | X_\lambda = X)}{E_x(\theta^t)}$$

where

$$\theta^t = \exp \left( -1/2 \int_0^t \frac{m^2}{q}(\tau) d\tau + \int_0^t \frac{m}{q}(\tau) [Z'(\tau) dZ(\tau)]_{12} \right) \quad (17)$$

with  $[Z'(\tau) dZ(\tau)]_{12} = [1, 0] Z'(\tau) dZ(\tau) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ .

If the density function of  $v_x$  exists and is denoted by  $p_{x_\lambda}(\cdot)$ , then it follows from Lemma 2 that the density function  $P_{x_\lambda}(\cdot | Z^t)$  of

$v_{xz}$  exists and can be expressed as follows:

$$p_{x_\lambda}(X | Z^t) = \frac{E_x(\theta^t | X_\lambda = X) p_{x_\lambda}(X)}{E_x(\theta^t)}$$

where  $\theta^t$  is defined by (17). Let  $x \in \mathbb{R}^1$  be defined by  $\exp Rx = X$  and  $-\pi \leq x < \pi$ . Then by simple calculations,

$$p_{x_\lambda}(X | Z^t) = E_x(\theta^t | x_\lambda = x + 2k\pi, k=1, 2, \dots) p_{x_\lambda}(X) / E_x(\theta^t)$$

$$= \sum_{k=-\infty}^{\infty} E_x(\theta^t | x(\lambda) = x + 2k\pi) p_{x_\lambda}^{(x+2k\pi)} / E_x(\theta^t),$$

where  $p_{x_\lambda}$  denotes the density function of  $x(\lambda)$ . This completes the

proof of the following theorem.

**Theorem 5:** Consider the observation process described by (15). If the density function  $p_{\underline{x}_\lambda}$  of  $X(\lambda)$  exists, then the conditional density function  $p_{\underline{x}_\lambda}(\cdot|Z^t)$  exists and can be expressed as follows

$$\begin{aligned} p_{\underline{x}_\lambda}(X|Z^t) &= \sum_{k=-\infty}^{\infty} p_{\underline{x}_\lambda}(x+2k\pi|Z^t) \\ &= \sum_{k=-\infty}^{\infty} \frac{E_x(\theta^t | x(\lambda) = x+2k\pi) p_{\underline{x}_\lambda}(x+2k\pi)}{E_x(\theta^t)} \end{aligned}$$

where  $\theta^t$  is defined by (17),  $p_{\underline{x}_\lambda}$  denotes the density function of  $x(\lambda)$  and  $x$  is determined by  $\exp Rx = X$  and the condition  $-\pi \leq x < \pi$ .

It is appropriate to remark that one can easily derive the stochastic partial differential equation for the conditional density  $p_{\underline{x}_\lambda}(X|Z^t)$  using Theorem 5 and the well-known equation ([31],[32]), for  $p_{\underline{x}_\lambda}(x+2k\pi|Z^t)$ ,  $-\infty < k < \infty$ . For economy of space, this equation will not be displayed. However we remark that when  $m(x,t)$  is periodic in  $x$  with period  $2\pi$ , the equation is in a form similar to the Stratonovich-Kushner equation with  $p_{\underline{x}_\lambda}$  replaced by  $p_{\underline{x}_\lambda}$ .

Using Theorem 5 and the well-known fact [30] that the smoothed and the predicted densities can be expressed explicitly in terms of filtering, we can easily obtain the following two corollaries.

**Corollary 1:** The conditional smooth density  $p_{\underline{x}_\lambda}(X|Z^t)$ , for  $t_0 \leq \lambda \leq t$ , may be expressed in terms of the conditional filtered density as follows:



$$p_{\tilde{x}_\lambda}(X|Z^t) = \sum_{k=-\infty}^{\infty} p_{x_\lambda}(x+2k\pi|Z^\lambda) \exp \left( \int_\lambda^t \frac{\alpha_s}{q(s)} dI_s - 1/2 \int_\lambda^t \frac{\alpha_s^2}{q(s)} ds \right)$$

where  $x$  is determined by  $\exp Rx = X$ ,  $-\pi \leq x \leq \pi$  and

$$dI_s = [Z'(s)dZ(s)]_{12} - \hat{m}(s)ds$$

$$\alpha_s = \hat{m}(s|x_\lambda = x) - \hat{m}(s)$$

$$\hat{m}(s) = E(m(s)|Z^s)$$

$$\hat{m}(s|x_\lambda = x) = E(m(s)|Z^s, x_\lambda = x).$$

**Corollary 2:** Let  $X$  be a Markov process with given transition density

$p_{\tilde{x}_\lambda}(X|x(t) = \xi)$ . The conditional predicted density  $p_{\tilde{x}_\lambda}(X|Z^t)$ , for  $t_0 \leq t \leq \lambda$  may be expressed in terms of the conditional filtered density as follows:

$$p_{\tilde{x}_\lambda}(X|Z^t) = \int_{-\infty}^{+\infty} p_{\tilde{x}_\lambda}(X|x(t) = \xi) p_{x_t}(\xi|Z^t) d\xi.$$

#### IV. 3. Optimal Estimation.

In the previous subsection, the conditional probability distributions were studied. A variety of estimation problems may be studied based on those conditional distributions, but some estimation problems on the circle can be solved directly by using results in vector-space estimation theory. In this subsection, the well-established linear optimal estimation theory will be used to deduce recursive equations for optimal estimation on  $S^1$  and thereby illustrate the approach.

The estimation problem with which we will mainly be concerned in this subsection is the following: Given a symmetric cost function  $\phi$  defined by (1), construct a  $2 \times 2$  orthogonal random matrix  $\hat{X}(\lambda | t)$  as a  $\mathcal{G}_1^t$ -measurable functional of  $Z^t$  such that for all  $\mathcal{G}_2$ -measurable  $2 \times 2$  orthogonal random matrices  $M$  one has the inequality

$$E(\phi(X(\lambda), \hat{X}(\lambda | t)) | Z^t) \leq E(\phi(X(\lambda), M) | Z^t), \quad (18)$$

in which  $\phi(X_1, X_2) \triangleq \phi(\theta)$ ,  $\theta$  being determined by  $\exp R\theta = X_1^{-1}X_2$  and the condition  $-\pi \leq \theta < \pi$  (i.e.,  $\theta$  is the angle between  $X_1$  and  $X_2$ ).

We have seen, at the beginning of this section, that a continuous random process  $X$  on  $S^1$  can be identified with a continuous random process  $x$  on  $R^1$  via the bijective mapping  $X = J(x)$ . We now construct a signal process  $X$  on  $S^1$  by injecting a linear diffusion  $x$  into  $S^1$ ,  $x$  satisfying

$$dx(t) = a(t)x(t) dt + b^{1/2}(t) dv(t), \quad x(0) = 0$$

where  $b(t) > 0$ ,  $\forall t \in T$ , and  $v$  is a standard Brownian motion, independent of the observational noise  $w$ . Applying the stochastic differentiation rule, we obtain the following stochastic differential equation for our signal process  $X = J(x)$ :

$$dX(t) = -1/2 b(t)X(t)dt + X(t)R\{a(t)[\int_0^t (\exp \int_s^t a(\tau)d\tau) b^{1/2}(s)dv(s)]dt + b^{1/2}(t)dv(t)\} \quad (19)$$

$$X(0) = I$$

where we note that  $x(t) = \int_0^t (\exp \int_0^\tau a(\tau)d\tau) b^{1/2}(s)dv(s)$ .

The observation process to be used in this subsection is taken to be  $Z$ , satisfying the stochastic differential equation:

$$dZ(t) = Z(t) \begin{bmatrix} \frac{-q(t)}{2} & c(t)x(t) \\ -c(t)x(t) & \frac{-q(t)}{2} \end{bmatrix} dt + Z(t) \begin{bmatrix} 0 & dw(t) \\ -dw(t) & 0 \end{bmatrix} \quad (20)$$

$$Z(0) = I.$$

As shown in Subsection IV.1,  $Z$  can be identified with  $z=J^{-1}(Z)$  satisfying

$$dz(t) = c(t)x(t)dt + q^{1/2}(t)dw(t)$$

$$z(0) = 0$$

Note that the equations for  $X$  and  $Z$  are each bilinear in form.

Moreover,  $z^t$  and  $Z^t$  generate the same  $\sigma$ -subfield  $A_z$  in  $(\Omega, A, P)$ . Hence  $E(x(\lambda) | A_z^t)$  is both a  $B_1^t$ -measurable functional  $f_1$  of  $z^t$  and a  $B_2^t$ -measurable functional  $f_2$  of  $Z^t$  with

$$f_2(Z^t) = f_1(J^{-1}(Z^t)). \quad (21)$$

Let  $\hat{x}_{\lambda|t}$  and  $\hat{x}(\lambda|t)$  denote  $f_1(z^t) = E(x(\lambda) | z^t)$  and  $f_2(Z^t) = E(x(\lambda) | Z^t)$  respectively.

We will first study the filtering problem, where  $\sigma = t$ . Then the Kalman-Bucy linear filtering theory yields immediately

$$d\hat{x}_{t|t} = a(t)\hat{x}_{t|t}dt + K(t)c(t)q^{-1}(t)(dz(t) - c(t)\hat{x}_{t|t}dt)$$

$$\hat{x}_{0|0} = 0$$

where  $K$  is the solution of the Riccati equation

$$\dot{K}(t) = 2a(t)K(t) - c^2(t)q^{-1}(t)K^2(t) + b(t)$$

$$K(0) = 0.$$



In view of (21), we obtain the following lemma, which not only leads to the solution of the above stated filtering problem but also applies directly to optimal frequency demodulation [6].

Lemma 3: Let the stochastic process (19) be the signal process and the stochastic process (20) be the observation process. Then the filtering equations are

$$\begin{aligned} d\hat{x}(t|t) = & a(t)\hat{x}(t|t)dt \\ & + K(t)c(t)q^{-1}(t) ([Z'(t)dZ(t)]_{12} - c(t)\hat{x}(t|t)dt) \end{aligned}$$

$$\hat{x}(0|0) = 0$$

$$\text{with } K(t) = 2 a(t)K(t) - c^2(t)q^{-1}(t)K^2(t) + b(t)$$

$$K(0) = 0$$

and the conditional probability density is given by

$$p_{x_t}(x|Z^t) = \frac{1}{\sqrt{2\pi} K(t)} \exp \left[ -\frac{1}{2K(t)} (x - \hat{x}(t|t))^2 \right].$$

In view of Theorem 5, we see that  $p_{\tilde{x}_t}(X|Z^t)$  is a folded normal density. By Theorem 2, it follows that  $p_{\tilde{x}_t}(X|Z^t)$  is unimodal with mode at  $\exp[\hat{x}(t|t)R]$  and is symmetric about it. We may now conclude from Theorem 1 that for a cost function defined by (1),

$$E(\phi(X(t), \exp[\hat{x}(t|t)R])|Z^t) \leq E(\phi(X(t), M)|Z^t)$$

for any  $A_Z$ -measurable  $2 \times 2$ -dimensional orthogonal random matrix  $M$ .

Since  $\exp[\hat{x}(t|t)R]$  is easily seen to be a  $B_1^t$ -measurable functional of  $Z^t$ , it follows that the optimal estimate of our signal process is

$$\hat{X}(t|t) = \exp[\hat{x}(t|t)R].$$

Differentiating this with respect to  $t$  yields

$$\begin{aligned} d\hat{X}(t|t) = & \frac{1}{2} K^2(t) c^2(t) q^{-1}(t) \hat{X}(t|t) dt + \hat{X}(t|t) R((a(t) - K(t) c^2(t) q^{-1}(t)) \\ & \cdot \left[ \int_0^t \left( \exp \int_s^t (a(\tau) - K(\tau) c^2(\tau) q^{-1}(\tau)) d\tau \right) K(s) c(s) q^{-1}(s) [Z'(s) dZ(s)]_{12} \right] dt \\ & + K(t) c(t) q^{-1}(t) [Z'(t) dZ(t)]_{12}. \end{aligned}$$

Summarizing what has been shown, we obtain the following theorem.

Theorem 6: If the signal process  $X$  and the observation process  $Z$  on  $S^1$  satisfy the following stochastic differential equations:

$$\begin{aligned} dX(t) = & \frac{1}{2} b(t) X(t) dt \\ & + X(t) R(a(t) \left[ \int_0^t \left( \exp \int_s^t a(\tau) d\tau \right) b^{1/2}(s) dv(s) dt \right] + b^{1/2}(t) dv(t)) \end{aligned}$$

$$X(0) = I$$

$$\begin{aligned} dZ(t) = & Z(t) \left[ \begin{array}{cc} -\frac{q(t)}{2} & c(t) \left[ \int_0^t X'(s) dX(s) \right]_{12} \\ -c(t) \left[ \int_0^t X'(s) dX(s) \right]_{12} & -\frac{q(t)}{2} \end{array} \right] dt \\ & + Z(t) \left[ \begin{array}{cc} 0 & dw(t) \\ -dw(t) & 0 \end{array} \right] \end{aligned}$$

$$Z(0) = I$$

where  $w$  and  $v$  are independent standard Brownian motions on  $R^1$ , then

the optimal estimate  $\hat{X}(t|t)$  in the sense of (18) satisfies the following stochastic differential equations:

$$\begin{aligned} d\hat{X}(t|t) = & -\frac{1}{2} K^2(t) c^2(t) q^{-1}(t) \hat{X}(t|t) dt + \\ & \hat{X}(t|t) R((a(t) - K(t) c^2(t) q^{-1}(t)) \left[ \int_0^t \left( \exp \int_s^t (a(\tau) - \right. \right. \\ & K(\tau) c^2(\tau) q^{-1}(\tau)) d\tau \cdot K(s) c(s) q^{-1}(s) [Z'(s) dZ(s)]_{12} \right] dt \\ & + K(t) c(t) q^{-1}(t) [Z'(t) dZ(t)]_{12} \end{aligned} \quad (22)$$

$$\dot{\hat{K}}(t) = 2a(t)K(t) - c^2(t)q^{-1}(t)K^2(t) + b(t) \quad (23)$$

$$K(0) = 0$$

The conditional probability density is given by

$$\begin{aligned} p_{\hat{x}|t}(x|z^t) &= \frac{1}{\sqrt{2\pi K(t)}} \sum_{k=-\infty}^{\infty} \exp \left[ -\frac{1}{2K(t)} (x + 2k\pi - \hat{x}(t|t))^2 \right] \\ &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \exp \left[ -\frac{k^2 K^2(t)}{2} \right] \cos k(x - \hat{x}(t|t)) \end{aligned}$$

where  $x$  is defined by  $\exp Rx = X$  and  $-\pi \leq x < \pi$ .

Some of the expected errors  $E(\phi(X(t), \hat{X}(t|t)))$  of the optimal estimate  $\hat{X}(t|t)$  can be obtained immediately from Theorem 4.

We note that the optimal filtering equations (22) and (23) are complex in form. Conceptually, however, the filtering procedure is quite simple: The observation process  $dZ$  first goes through a nonlinear transformer. The transformed process  $[Z'dZ]_{12}$  then goes through a Kalman-Bucy linear filter. Then we inject the filtered process  $\hat{x}(t|t)$  into  $S^1$  via the injection mapping  $J$ . The output  $\hat{X}(t|t)$  of the nonlinear injector is the desired estimate.

The same approach can be used to solve the smoothing and prediction problems. The solution to the prediction problem is trivial and hence omitted here. For the smoothing problem, we first recall [33] that for  $0 \leq \lambda \leq t$ ,

$$\begin{aligned} \hat{x}_{\lambda|t} &= \hat{x}_{\lambda|\lambda} + K(\lambda) \int_{\lambda}^t \left( \exp \int_{\lambda}^s (a(\tau) - K(\tau)c^2(\tau)q^{-1}(\tau)) d\tau \right) c(s) \\ &\quad \cdot q^{-1}(s) (dz(s) - c(s)\hat{x}_s|_s ds). \end{aligned}$$

By (21), it follows that



$$\hat{x}(\lambda|t) = \hat{x}(\lambda|\lambda) + K(\lambda) \int_{\lambda}^t \left( \exp \int_{\lambda}^s (a(\tau) - K(\tau)c^2(\tau)q^{-1}(\tau))d\tau \right. \\ \left. \cdot c(s)q^{-1}(s)([Z'(s)dZ(s)]_{12} - c(s)\hat{x}(s|s)ds) \right). \quad (24)$$

We note that the conditional probability distribution of  $x(\lambda)$  given  $Z^t$  is Gaussian. From Theorem 5, it follows that  $p_{\hat{x}_\lambda}(X|Z^t)$  is a folded-Gaussian density and hence unimodal. As in the filtering case,

$$\hat{X}(\lambda|t) = \exp(\hat{x}(\lambda|t)R). \quad (25)$$

Substituting (24) into (25) thus yields

$$\hat{X}(\lambda|t) = \hat{X}(\lambda|\lambda) \exp \left\{ RK(\lambda) \int_{\lambda}^t \left( \exp \int_{\lambda}^s (a(\tau) - K(\tau)c^2(\tau)q^{-1}(\tau))d\tau \right) \right. \\ \left. c(s)q^{-1}(s) ([Z'(s)dZ(s)]_{12} - c(s) \int_0^s [\hat{X}'(\tau|\tau)d\hat{X}(\tau|\tau)]_{12} ds) \right\}$$

where we have used the identity  $\hat{x}(s|s) = \int_0^s [\hat{X}'(\tau|\tau)dX(\tau|\tau)]_{12}$ .

Summarizing what has been shown, we obtain the following theorem.

**Theorem 7:** If the signal process and the observation process are the same as in Theorem 5, then the optimal estimate,  $\hat{X}(\lambda|t)$ ,

$0 \leq \lambda \leq t$ , in the sense of (52), is given by

$$\hat{X}(\lambda|t) = \hat{X}(\lambda|\lambda) \exp \left\{ RK(\lambda) \int_{\lambda}^t \left( \exp \int_{\lambda}^s (a(\tau) - K(\tau)c^2(\tau)q^{-1}(\tau))d\tau \right) \right. \\ \left. c(s)q^{-1}(s) ([Z'(s)dZ(s)]_{12} - c(s) \int_0^s [\hat{X}'(\tau|\tau)d\hat{X}(\tau|\tau)]_{12} ds) \right\},$$

where  $\hat{X}(\tau|\tau)$ ,  $K(\tau)$  can be obtained from (22) and (23).

The conditional probability density of  $X(\lambda)$  given  $Z^t$ , the expected errors  $E(\phi(X(\lambda), \hat{X}(\lambda|t)))$ , the stochastic equations for  $\hat{X}(\lambda|t)$  for fixed-point smoothing, fixed-lag smoothing, and fixed interval smoothing can all be easily obtained by straightforward computations which are left to the interested reader.

#### IV. 4. Random Initial State.

In the previous subsections, the initial state of the signal process  $X$  has been assumed to be  $X(0) = I$ , the identity matrix. This is obviously not a practical assumption in some applications. In this subsection we will consider the case in which the initial state is a random variable. We will denote the signal process by  $Y$  in this subsection, and assume that  $Y(0) = Y_0$  is a random variable independent of the observational noise  $w$ .

We observe that the input to the observation process (15) at time  $t$  is not the dynamical state of the signal. It is the angle that the rotational process represented by the signal has swept over the time interval  $[0, t]$ . Taking this viewpoint, our present problem can be solved through the previous ideas with some modification.

Let  $y(t)$  denote the angle through which the signal  $Y$  has swept during  $[0, t]$ . It is easily seen that

$$y(t) = \left[ \int_0^t Y'(s) dY(s) \right]_{12}$$

Define a rotational process  $X$  by

$$X(t) = Y_0^{-1} Y(t)$$

Then  $X(0) = I$  and, as before, we may define

$$x(t) = (J^{-1}(X))(t) = \left[ \int_0^t X'(s) dX(s) \right]_{12}.$$

Note that  $x(t) = y(t)$ . In other words, the angles swept by  $X$  and by  $Y$  over  $[0, t]$  are the same. Hence (15) can also be used as the observation process for our present problem. The conditional distribution of  $X(\lambda)$ , given observation  $Z^t$  of the form given in (15), can be determined by application of the previous results.

We note that  $Y_0$  and  $X(\lambda)$  are conditionally independent given  $Z^t$ . If the distribution of  $Y_0$  and the conditional distribution of  $X(\lambda)$  given  $Z^t$  are both folded normal, then the following lemma easily leads to the conclusion that  $Y(\lambda|t)$ , the optimal estimate of  $Y(\lambda)$  given  $Z^t$ , is equal to  $\hat{Y}_0 \hat{X}(\lambda|t)$ . Here  $\hat{Y}_0$  is the mode of the distribution of  $Y_0$  and  $\hat{X}(\lambda|t)$  is the mode of the conditional distribution of  $X(\lambda)$  given  $Z^t$ .

Lemma 4: Let  $A$  and  $B$  be two independent random  $2 \times 2$  orthogonal matrices each of which has a folded normal distribution with modes  $\hat{A}$  and  $\hat{B}$  respectively. Then  $AB$  is a random  $2 \times 2$  orthogonal matrix which has a folded normal distribution with mode  $\hat{A}\hat{B}$ .

Proof. It is easily seen that there exist unique real-valued normal random variables  $a$  and  $b$  such that  $Ea, Eb$  are in  $[-\pi, \pi)$  with  $A = \exp Ra$ , and  $B = \exp Rb$ . Then  $AB = \exp R(a+b)$ . Obviously  $a+b$  is a normal random variable. Hence  $AB$  is folded normal and the mode of  $AB$  is  $\exp[RE(a+b)] = \exp[RE(a)] \cdot \exp[RE(b)] = \hat{A}\hat{B}$ .

#### IV. 5. Multichannel Estimation.

The results of the previous subsections can be extended to the large class of problems involving processes evolving on abelian Lie groups. It is well known [34] that a given connected abelian Lie group  $G$  is isomorphic to the direct product of a number of copies of the circle and a number of copies of the real line, i.e.

$$G \approx \mathbb{R}^n \times (S^1)^m$$



where  $(S^1)^m$  is usually called a "torus." The diffusion processes on this type of space have been used to model some interesting satellite and pendulum systems in [54]. Analogous to (14), a bijective mapping  $J_{nm} : (C_1^S)^{n+m} \rightarrow (C_1^S)^n \times (C_2^S)^m$  is defined by

$$(J_{nm}(a))(t) = [a_1(t), \dots, a_n(t), (J(a_{n+1}))(t), \dots, (J(a_{n+m}))(t)]$$

for  $a \in (C_1^S)^{n+m}$ ,  $a_i$  being the  $i$ th component of  $a$ . Thus a continuous random signal process on  $G$  which is described by an  $A$ -measurable function  $X: \Omega \rightarrow (C_1^S)^n \times (C_2^S)^m$  corresponds to a unique continuous random signal process on  $R^{n+m}$  which is described by an  $A$ -measurable function  $x: \Omega \rightarrow (C_1^S)^{n+m}$  such that

$$X(t) = (J_{nm}(x))(t), \quad t \in [0, S].$$

The mathematical model for the sensor can be obtained by first using  $J_{nm}$  to inject the following vector random differential equation into  $R^n \times (S^1)^m$

$$\begin{aligned} dz(t) &= m(x(t), t)dt + dv(t) \\ z(0) &= 0 \end{aligned} \tag{26}$$

and then differentiating  $Z(t) = (J_{nm}(z))(t)$  by the stochastic differentiation rule to obtain a set of stochastic differential equations. The first  $n$  of these equations are the same as the first  $n$  equations of (26) and the last  $m$  equations are bilinear  $2 \times 2$  matrix differential equations having the form (15). This calculation is straightforward and so we will not display the resulting equations. Because of the bijective property of  $J_{nm}$ , it is clear that the estimation analysis in the previous subsections can be easily generalized to this general abelian case with little modification. For the special case in which  $x$  is a linear diffusion and

$m(x(t), t)$  is a linear function of  $x(t)$ , what has been shown is that the applicability of the celebrated Kalman-Bucy filter includes estimation on abelian Lie groups.

#### V. DISCRETE-TIME ESTIMATION ON COMPACT LIE GROUPS

The results of Section III can easily be generalized to problems on compact non-abelian Lie groups by introducing a similar exponential Fourier density (EFD) on the group. This density is obtained by using a sequence of irreducible unitary representations which form a complete orthogonal system on the compact group. It can be shown that a continuous density function on the group can be approximated as closely as we wish in the space of square integrable functions by such an EFD.

As with the circle case a consequence of the group structure is that the class of EFD's of a certain finite order on the compact Lie group is closed under the operation of taking conditional distributions. It will become clear in the sequel that it is exactly this closure property of the EFD's that yields simple estimation schemes in which the sequential conditional densities are updated by recursively revising a fixed finite number of parameters.

In order to illustrate how the conditional density can be used to calculate the optimal estimate on the group, a rigid body attitude estimation problem is solved as an example. The error criterion, the optimal estimate, and the estimation error with respect to the criterion will be discussed for a given probability distribution.

# V. I. Compact Lie Groups and Their Matrix Representations.

We begin by summarizing some definitions and preliminary results to be used in this section. The reader is referred to [35]-[37] for details.

Definition. A differential manifold  $M$  of dimension  $n$  is a Hausdorff topological space with the following properties: (a) For every element  $m \in M$  there are an open set  $U$  containing it and a homeomorphism  $y: U \rightarrow V \subset \mathbb{R}^n$ , called a chart. The set  $V$  is called a parameter domain. The components of vector  $y(m)$  are called the coordinates of  $m$ , (b) For any two charts,  $y_1$  and  $y_2$ , defined on  $U_1$  and  $U_2$ , the composition  $y_2 \circ y_1^{-1}$  defined on  $y_1(U_1 \cap U_2)$  is smooth (i.e., infinitely differentiable).

Definition. A Lie group  $G$  is both a differential manifold and a group, which is closed and connected, such that the group operations are smooth in coordinates. If the group is covered by finite number of bounded parameters domains through their charts, then the group is said to be compact.

Definition. An  $m \times m$  matrix representation of a Lie group  $G$  is a subgroup  $\Gamma$  of the nonsingular  $m \times m$  matrices together with a homomorphic smooth mapping  $D$  of  $G$  onto  $\Gamma$ . That is, for each  $a, b \in G$ , there is an element  $D(a) \in \Gamma$  such that (a)  $D(a)D(b) = D(ab)$ , (b)  $D(e) = I$ , and (c)  $D(a^{-1}) = [D(a)]^{-1}$ . We write  $\dim D = m$ . The representation is said to be unitary if each matrix in  $\Gamma$  is unitary. Two such representations  $D^1$  and  $D^2$  are called equivalent if there is a nonsingular  $m \times m$  matrix  $\psi$  such that  $\psi D^1(a) = D^2(a) \psi$  for each  $a \in G$ . A reducible representation is one that is equivalent to the block form,



$$D(a) = \begin{bmatrix} D^1(a) & C(a) \\ 0 & D^2(a) \end{bmatrix},$$

where  $D^1$  and  $D^2$  can be shown to constitute representations. If a representation is equivalent to such a block form with  $C = 0$ , it is called completely reducible. It can be shown that a reducible unitary representation must be completely reducible.

Definition. We delete from some of the parameter domains their intersection with others so that the points of the resulting domains are in 1-1 correspondence with the group elements. Then the integral  $\int f(x)w(x)dx$  of the function  $f$  with respect to the weight function  $w$  is well defined. It can be shown that a weight function  $w$ , unique except for a normalizing factor, can be found such that this integral is left invariant, e.e.,  $\int f(p)w(p)dp = \int f(ap)w(p)dp$  for any continuous function  $f$  and any group element  $a$ . On a compact Lie group the integral is also right invariant and is written as  $\int f(g)dg$ .

Theorem 8. Let  $D^1(a), D^2(a), \dots$ , be a family of inequivalent irreducible unitary representations of a compact Lie group. The matrix elements  $D_{ij}^k$  of these representations satisfy the orthogonality relations

$$\int D_{im}^{\ell}(g) D_{jn}^k(g)^* dg = (\int dg / \dim(D^{\ell})) \delta_{\ell k} \delta_{ij} \delta_{mn}.$$

Theorem 9. (Peter-Weyl) A continuous function on a compact Lie group can be uniformly approximated by a linear combination of the matrix elements  $D_{ij}^k$  of the unitary irreducible representations of the group.

V. 2. Exponential Fourier Densities on a Compact Lie Group.

Let us denote by  $D^1, D^2, \dots$ , a collection of irreducible, inequivalent, and unitary matrix representations of a compact Lie group  $G$ , which are of dimensions  $n_1, n_2, \dots$  respectively. We define an exponential Fourier density of order  $N$ , to be denoted by  $EFD(N)$ , on  $G$  as a probability density of the form

$$p(a) = \exp \sum_{\ell=0}^N \sum_{i,j=1}^{n_\ell} a_{ij}^\ell D_{ij}^\ell(a) \triangleq$$

$$\exp \left( \sum_{\ell=1}^N \sum_{i,j=1}^{n_\ell} a_{ij}^\ell D_{ij}^\ell(a) + a_{00}^0 \right)$$

where  $a_{00}^0$  is a normalizing constant and all other coefficients  $a_{ij}^\ell$  are arbitrary complex numbers. The double summation notation above will be abbreviated by  $\Sigma$ . The norm of a function  $f$  in  $L^2(G)$  will be denoted by  $||f|| \triangleq \left( \int |f|^2 dg \right)^{1/2}$ .

Theorem 10. Let  $p$  be a probability density on a compact Lie group  $G$ . Assume that  $p$  is continuous. Then for any given positive number  $\epsilon$ , there exists an EFD,  $p_N = \exp \Sigma a_{ij}^\ell D_{ij}^\ell$ , such that  $||p - p_N|| \leq \epsilon$ .  
 Proof: Assume that

$$\inf\{p(x) : x \in G\} = c > 0. \quad (27)$$

This assumption will be removed later. We note that  $f(x) = \ln p(x)$  is then well defined and also continuous on  $G$ .

Since  $G$  is compact, in view of the Peter-Weyl Theorem,

for any  $0 < \delta < 1$  there is a linear combination of  $D_{mn}^1$ , say  $f_\delta = \sum a_{mn}^\delta(\delta) D_{mn}^\delta$ , such that  $\|f_\delta - f\|_\infty < \delta$ . It follows that  $\|f_\delta\|_\infty < 1 + \|f\|_\infty =: M$ .

Define a function  $g: \mathbb{R}^1 \rightarrow \mathbb{R}^1$  by

$$g(x) = \begin{cases} \exp x, & x \leq M \\ \exp M, & x > M \end{cases} \quad g(x) = \exp \min \{x, M\}$$

and an operator  $\bar{g}$  on the set of real functions on  $G$  by  $(\bar{g}u)(x) = g(u(x))$ .

It is obvious that  $g$  satisfies the Caratheodory conditions [38,p.20] and  $\bar{g}$  transforms every function in  $L^2(G)$  into a function in  $L^2(G)$ .

By Theorem 2.1 of [38,p.22], the operator  $\bar{g}$  is continuous. Hence given any  $\epsilon > 0$ , there exists a  $\delta > 0$  such that if  $\|f_\delta - f\| < \delta$ , then  $\|\bar{g}f_\delta - \bar{g}f\| < \epsilon$ . Then  $\|\exp f_\delta - p\| = \|\bar{g}f_\delta - \bar{g}f\| < \epsilon$ .

Now let us remove the assumption (27) and assume that  $\inf\{p(x): x \in G\} = 0$ . Let  $\epsilon$  be an arbitrary positive number. Set  $\epsilon_1 \triangleq \epsilon_2 = \epsilon/2$  and  $p_1(x) \triangleq p(x) + \epsilon_1/V$ , where  $V$  is the volume of  $G$ . The function  $p_1$  satisfies (27). Hence there is an EFD(N),  $\exp f$ , such that  $\|\exp f - p_1\| \leq \epsilon$ . By the Minkowski inequality,

$$\|\exp f - p\| \leq \|\exp f - p_1\| + \|p_1 - p\| \leq \epsilon_2 + \epsilon_2 = \epsilon. \quad (28)$$



So far we have shown that for any  $\epsilon > 0$ , there is an EFD(N),  $\exp f$ , such that  $||\exp f - p|| \leq \epsilon$ . We note that  $\exp f$  is not necessarily a probability density function. In the following it will be shown that the exponential Fourier density obtained by dividing  $\exp f$  by  $\int \exp f(g) dg$  can be used to approximate  $p$  as closely as desired.

$$\begin{aligned} \text{Let } \hat{p}_1 &\triangleq \exp f \text{ and } \hat{p} \triangleq \hat{p}_1 / \int \hat{p}_1(g) dg. \text{ Then} \\ ||\hat{p}_1 - \hat{p}|| &= \left[ \int \left( \frac{\hat{p}_1(g)}{\int \hat{p}_1(g) dg} - \hat{p}_1(g) \right)^2 dg \right]^{1/2} \\ &= |1 / \int \hat{p}_1(g) dg - 1| \left( \int \hat{p}_1^2(g) dg \right)^{1/2} \\ &\leq |1 / \int \hat{p}_1(g) dg - 1| \sqrt{V} \exp M. \end{aligned} \quad (29)$$

By the Hölder inequality,

$$\begin{aligned} |1 - \int \hat{p}_1(g) dg| &= |\int (p(g) - \hat{p}_1(g)) dg| \\ &\leq \int |p(g) - \hat{p}_1(g)| dg \\ &\leq \left( \int (p(g) - \hat{p}_1(g))^2 dg \right)^{1/2} \sqrt{V} \\ &\leq \sqrt{V} \epsilon. \end{aligned}$$

Hence,  $\int \hat{p}_1(g) dg \geq 1 - \sqrt{V} \epsilon$ . Substituting these two inequalities into (29), we have  $||\hat{p}_1 - \hat{p}|| \leq (V\epsilon / (1 - \sqrt{V}\epsilon)) \exp M$ . By the Minkowski inequality, (28), and (29),

$$||\hat{p}_1 - p|| \leq ||\hat{p}_1 - \hat{p}|| + ||\hat{p} - p|| \leq \frac{V\epsilon}{1 - \sqrt{V}\epsilon} \exp M + \epsilon =: \hat{\epsilon}.$$

We observe that  $\hat{\epsilon}$  can be made arbitrarily small by setting  $\epsilon$  sufficiently small. This observation completes the proof of the theorem.

### V. 3. Estimation for Processes with Noise on the Lie Group.

Suppose that the signal process  $s_k$  and the measurement process  $m_k$  both evolve on a compact Lie group  $G$  and are related by

$$m_k = v_k \circ s_k \quad (30)$$

where  $v_k$  is a noise process, also evolving on  $G$  and  $\circ$  denotes the group operation. Our reason for writing  $v$  to the left of  $s$  is to be consistent with the corresponding matrix equation obtained by the use of the orthogonal matrix representation: the matrix  $S$  representing  $s$  is premultiplied by the matrix  $V$  representing  $v$  to obtain  $M = VS$ .

We now consider a signal process  $s_k$  which is governed by the equation

$$s_{k+1} = w_k \circ s_k \quad (31)$$

where  $w_k$  is a sequence of known elements on  $G$ . If  $s_0$  is a random variable taking values on  $G$ , an interesting estimation problem is to find an effective way to recursively compute the conditional density of  $s_k$  given the set of measurements,  $m^k \triangleq \{m_1, \dots, m_k\}$ ,  $k=1, 2, \dots$

The EFD's introduced previously are ideal to use in solving this problem on many compact Lie groups such as the three dimensional rotation group,  $SO(3)$ . However, for a reason to be discussed later, it is more convenient to include the complex conjugates  $D_{ij}^{\ell*}$  of the harmonic functions  $D_{ij}^{\ell}$  in the EFD(N). Thus an EFD(N) in this subsection will be a density function in the form

$$p(a) = \exp \sum (a_{ij}^{\ell} D_{ij}^{\ell}(a) + b_{ij}^{\ell} D_{ij}^{\ell*}(a)).$$

Suppose  $s_0$  and  $v_k$  have EFD(N)'s (if they have different orders, we can let N be the maximum order and, by inserting zero coefficients, make all densities of order N) which are described, respectively, by

$$p(s_0) = \exp \sum (a_{mn}^{l0} D_{mn}^l(s_0) + b_{mn}^l D_{mn}^{l*}(s_0)) \quad (32)$$

$$p(v_k) = \exp \sum (\hat{a}_{mn}^{lk} D_{mn}^l(v_k) + \hat{b}_{mn}^{lk} D_{mn}^{l*}(v_k)) \quad (33)$$

We claim that if the conditional probability densities,  $p(s_k | m^k)$ ,  $k=1,2,\dots$ , are all EFD(N)'s, then we need only keep track of a fixed finite number,  $\sum_{l=1}^N n_l^2$ , of parameters for

updating the conditional densities. The proof is by mathematical induction.

For  $k=0$ ,  $p(s_k | m^k)$  is obviously an EFD(N), as  $p(s_0 | m^0) = p(s_0)$ . Let us assume that the conditional density  $p(s_{k-1} | m^{k-1})$  is an EFD(N), denoted by

$$p(s_{k-1} | m^{k-1}) = \exp \sum (a_{mn}^{1l,k-1} D_{mn}^l(s_{k-1}) + a_{mn}^{2l,k-1} D_{mn}^{l*}(s_{k-1})) \quad (34)$$

We will now show that  $p(s_k | m^k)$  is also an EFD(N) and at the same time exhibit a recursive formula for the Fourier coefficients

$$a_{mn}^{1lk} \quad \text{and} \quad a_{mn}^{2lk}.$$

From (30), (31) and the group property of G,  $v_k$  and  $s_{k-1}$  can be expressed as



$$v_k = m_k \circ s_k^{-1} \text{ and } s_{k-1} = w_{k-1}^{-1} \circ s_k$$

Thus, using (34),  $p(s_k | m^{k-1})$  is an EFD(N):

$$\begin{aligned} p(s_k | m^{k-1}) &= \exp \Sigma [a_{mn}^{1\ell, k-1} D_{mn}^{\ell}(w_{k-1}^{-1} \circ s_k) + a_{mn}^{2\ell, k-1} D_{mn}^{\ell*}(w_{k-1}^{-1} \circ s_k)] \\ &= \exp \Sigma [a_{mn}^{1\ell, k-1} \sum_{i=1}^n D_{mi}^{\ell}(w_{k-1}^{-1}) D_{in}^{\ell}(s_k) \\ &\quad + a_{mn}^{2\ell, k-1} \sum_{i=1}^n D_{mi}^{\ell*}(w_{k-1}^{-1}) D_{in}^{\ell*}(s_k)] \\ &= \exp \Sigma \{ [\sum_{j=1}^n a_{jn}^{1\ell, k-1} D_{jm}^{\ell}(w_{k-1}^{-1})] D_{mn}^{\ell}(s_k) + \\ &\quad [\sum_{j=1}^n a_{jn}^{2\ell, k-1} D_{jm}^{\ell*}(w_{k-1}^{-1})] D_{mn}^{\ell*}(s_k) \} \end{aligned} \quad (35)$$

The second equality holds because  $D$  is a matrix group representation so  $D^{\ell}(g_1 \circ g_2) = D^{\ell}(g_1) D^{\ell}(g_2)$ .

The following calculation shows that  $p(m_k | s_k)$  is also an EFD(N):

$$\begin{aligned} p(m_k | s_k) &= \exp \Sigma \{ b_{mn}^{1\ell k} D_{mn}^{\ell}(m_k \circ s_k^{-1}) + b_{mn}^{2\ell k} D_{mn}^{\ell*}(m_k \circ s_k^{-1}) \} \\ &= \exp \Sigma \{ [\sum_{j=1}^n b_{ju}^{1\ell k} D_{jm}^{\ell}(m_k)] D_{mn}^{\ell}(s_k^{-1}) + \\ &\quad [\sum_{j=1}^n b_{ju}^{2\ell k} D_{jm}^{\ell*}(m_k)] D_{mn}^{\ell*}(s_k^{-1}) \} \\ &= \exp \Sigma \{ [\sum_{j=1}^n b_{jn}^{1\ell k} D_{jm}^{\ell*}(m_k)] D_{nm}^{\ell*}(s_k) + \\ &\quad [\sum_{j=1}^n b_{jn}^{2\ell k} D_{jm}^{\ell*}(m_k)] D_{nm}^{\ell}(s_k) \} \end{aligned} \quad (36)$$

The last equality holds because  $D$  is a unitary matrix representation so  $D^\ell(g^{-1}) = [D^{\ell*}(g)]^*$ .

We note that the complex conjugates  $D_{mn}^{\ell*}$  are included in the EFD(N)'s in this subsection just to ensure that the above expression be an EFD(N). On many compact Lie groups, the complex conjugates  $D_{mn}^{\ell*}$  are unnecessary. For instance, on the three dimensional rotation group we have  $D_{mn}^\ell(g^{-1}) = (-1)^{m+n} D_{-n, -m}^\ell(g)$ ,  $m, n = -\ell, -\ell+1, \dots, \ell$ , where the complex conjugates are avoided.

Substituting (35) and (36) into the Bayes Rule, we obtain

$$\begin{aligned} p(s_k | m^k) &= c_k p(s_k | m^{k-1}) p(m_k | s_k) \\ &= c_k \exp \left\{ \sum_{j=1}^n \sum_{\ell} \left[ a_{jn}^{1\ell, k-1} D_{jm}^\ell(w_{k-1}^{-1}) + b_{jm}^{2\ell k} D_{jn}^{\ell*}(m_k) \right] D_{mn}^\ell(s_k) \right. \\ &\quad \left. + \sum_{j=1}^n \sum_{\ell} \left[ a_{jn}^{2\ell, k-1} D_{jm}^{\ell*}(w_{k-1}^{-1}) + b_{jm}^{1\ell k} D_{jn}^\ell(m_k) \right] D_{mn}^{\ell*}(s_k) \right\}, \end{aligned}$$

which is an EFD(N). This completes the proof of the following:

**Theorem 11.** Let the signal and the measurement processes,  $s_k$  and  $m_k$ , on a compact Lie group  $G$  be governed by

$$s_{k+1} = w_k \circ s_k$$

$$m_k = v_k \circ s_k$$

Here  $w_k$  is a sequence of known elements on  $G$  and  $v_k$ , the measurement noise process, is a sequence of independent random variables taking values on  $G$ . Suppose the probability densities of  $s_0$  and  $v_k$  are EFD(N)'s described by (32) and (33). Then for  $k=1, 2, \dots$ , the conditional density  $p(s_k | m^k)$  is an EFD(N) of the form

$$p(s_k | m^k) = \exp \Sigma (a_{mn}^{1\ell k} D_{mn}^{\ell}(s_k) + a_{mn}^{2\ell k} D_{mn}^{\ell*}(s_k))$$

The coefficients  $a_{mn}^{1\ell k}$  and  $a_{mn}^{2\ell k}$  are determined recursively by the formulas

$$a_{mn}^{1\ell k} = \sum_{j=1}^{n\ell} [a_{jn}^{1\ell, k-1} D_{jm}^{\ell}(w_{k-1}^{-1}) + b_{jm}^{2\ell k} D_{jn}^{\ell*}(m_k)]$$

$$a_{mn}^{2\ell k} = \sum_{j=1}^{n\ell} [a_{jn}^{2\ell, k-1} D_{jm}^{\ell*}(w_{k-1}^{-1}) + b_{jm}^{1\ell k} D_{jn}^{\ell}(m_k)]$$

and  $a_{oo}^{ok}$  is a normalizing constant.

#### V. 4. Estimation for Processes with Additive Noise.

In this subsection we will consider another model for which the estimation problem can be solved using EFD(N)'s. Suppose that the signal process  $s_k$  evolves on a compact Lie group G according to the equation (31) and it is observed with additive noise  $v_k$  through the p-dimensional vector-valued measurement process  $m_k$ ,

$$m_k = h(s_k) + v_k$$

Here h is a given square-integrable, p-dimensional vector-valued function on G, and  $v_k$  is a sequence of p-dimensional independent Gaussian vectors, each having zero mean and with covariance matrices  $E(v_k v_k') = R_k$ .

The completeness property of the functions  $\{D_{mn}^{\ell}\}$  assures us that, for any  $\epsilon > 0$  and for each component  $h^j$  of the function h, there exists an integer  $M_j$  and coefficients  $h_{mn}^{\ell j}$  such that



$$\|h^j(s) - \sum_{\ell=1}^{M_j} \sum_{m,n=1}^n h_{mn}^{\ell j} D_{mn}^{\ell}(s)\|_2 < \epsilon, \quad j=1, 2, \dots, p.$$

Let  $M = \max_j M_j$  and denote by  $h_M(s)$  the  $p$ -dimensional vector whose

$$j\text{th component is } \sum_{\ell=1}^M \sum_{m,n=1}^n h_m^{\ell j} D_{mn}^{\ell}(s) \triangleq \sum_{\ell=1}^{M_j} \sum_{m,n=1}^n h_m^{\ell j} D_{mn}^{\ell}(s),$$

with  $h_m^{\ell j} \triangleq 0$  for  $\ell > M_j$ . For abbreviation, we will denote the double summation notation by  $\Sigma$ .

Since the function  $h$  is a mathematical description which is necessarily an approximation of the physical phenomenon that it describes, we may as well use the equation  $m_k = h_M(s_k) + v_k$  to represent the observation of the signal  $s_k$ .

Each noise vector has density

$$p(v_k) = (2\pi)^{-p/2} (\det R_k)^{-1/2} \exp \left[ -\frac{1}{2} \sum_{i,j=1}^p R_k^{ij} v_k^i v_k^j \right]$$

where  $R_k$  has components  $R_k^{ij}$  and  $v_k$  has components  $v_k^i$ . By substituting  $m_k - h_M(s_k)$  for  $v_k$  we obtain

$$\begin{aligned} p(m_k | s_k) &= (2\pi)^{-p/2} (\det R_k)^{-1/2} \\ &\exp \left\{ -\frac{1}{2} \sum_{i,j=1}^p R_k^{ij} [m_k^i - \Sigma h_{mn}^{\ell i} D_{mn}^{\ell}(s_k)] \times [m_k^j - \Sigma h_{mn}^{\ell j} D_{mn}^{\ell}(s_k)] \right\} \\ &= (2\pi)^{-p/2} (\det R_k)^{-1/2} \exp \{ C_0 + \Sigma C_{mn}^{\ell} D_{mn}^{\ell}(s_k) \\ &+ \sum_{\ell, \ell'=1}^M \sum_{m,n=1}^n \sum_{m',n'=1}^n C^{\ell \ell'}(m,n,m',n') D_{mn}^{\ell}(s_k) D_{m'n'}^{\ell'}(s_k) \} \end{aligned}$$

Here,

$$C_0 = -1/2 \sum_{j=1}^P R_k^{ij} m_k^i m_k^j$$

$$C_{mn}^{\ell} = 1/2 \sum_{k,j=1}^P R_k^{ij} [m_k^j h_{mn}^{\ell j} + m_k^i h_{mn}^{\ell i}]$$

$$C^{\ell \ell'}(m,n,m'n') = -1/2 \sum_{i,j=1}^P R_k^{ij} h_{mn}^{\ell i} h_{m'n'}^{\ell' j}.$$

We note that if the product function,  $D_{mn}^{\ell}(s_k) D_{m'n'}^{\ell'}(s_k)$ ,

can be expressed as a linear combination of finitely many harmonic functions,  $D_{mn}^{\ell}(s_k)$ , on  $G$ , then  $p(m_k | s_k)$  is an EFD of finite dimension. Fortunately, this is indeed the case. The product function  $D_{mn}^{\ell} D_{m'n'}^{\ell'}$  is a component of the direct product,  $D^{\ell} \times D^{\ell'}$ , of  $D$  and  $D'$ , which is itself a representation of  $G$  [35,p.79]. As every finite dimensional representation of a compact Lie group is equivalent to the direct sum of a finite number of irreducible unitary representations [36, p.333], the component  $D_{mn}^{\ell} D_{m'n'}^{\ell'}$  of the finite dimensional representation  $D^{\ell} \times D^{\ell'}$  is indeed a linear combination of finitely many  $D_{mn}^{\ell}$ 's, which we write

$$D_{mn}^{\ell} D_{m'n'}^{\ell'} = \sum_{i=1}^M \sum_{j,k=1}^n \lambda_{\ell'm'n'}^{\ell mn} (i,j,k) D_{jk}^i,$$

The  $\lambda$ 's are constants and  $M$  is the maximum superscript of all the irreducible unitary representations which appear in the above mentioned direct sum.

It was shown in (35) that if  $p(s_{k-1} | m^{k-1})$  is an EFD( $N_{k-1}$ ), then  $p(s_k | m^{k-1})$  is also an EFD( $N_{k-1}$ ). Therefore if  $p(s_{k-1} | m^{k-1})$  is an EFD( $N_{k-1}$ ), the conditional density  $p(s_k | m^k)$ , which is equal to  $c_k p(s_k | m^{k-1}) p(m_k | s_k)$ , is an EFD( $\max\{N_{k-1}, M\}$ ). Thus if  $p(s_0)$  is an EFD( $N$ ) given by (31), then  $p(s_k | m^k)$  will also be an EFD( $\max\{M, N\}$ )

for all  $k = 1, 2, \dots$ . The recursive formulas for updating the coefficients  $a_{mn}^{\ell k}$  can be easily obtained by straightforward but tedious calculations which are omitted here.

As remarked above, the determination of  $\lambda_{\ell' m' n'}^{\ell mn}$  ( $i, j, k$ ) depends on the decomposition of the direct product  $D^{\ell} \times D^{\ell'}$  of irreducible representations  $D^{\ell}$  and  $D^{\ell'}$ . Such a decomposition is not always easy but, fortunately, such decompositions have been thoroughly studied and documented for many special groups including  $SO(3)$ . The interested reader is referred to [40, p. 80] and [39, p. 155] for further discussions and references.

#### V. 5. An Example - Orientation Estimation of a Rigid Body Rotation.

The state space of a rigid body rotation is the three dimensional rotation group denoted by  $SO(3)$ . A common way to parametrize this group is to use the triple of Euler Angles  $(\phi, \theta, \psi)$ ,  $0 \leq \phi < 2\pi$ ,  $0 \leq \theta < \pi$ ,  $0 \leq \psi < 2\pi$ . Thus, each element of  $SO(3)$  is expressed uniquely as the result of a sequence of rotations through these angles about the  $z$  -  $x$  - and  $z$  - axes.

We will use a sequence of finite dimensional unitary representations  $\{D^{\ell}(\phi, \theta, \psi), \ell = 0, 1, \dots\}$  attributed to E. P. Wigner. The components are described in [35. p 144] by :



$$D_{mn}^{\ell}(\phi, \theta, \psi) = i^{m-n} e^{-im\phi} d_{mn}^{\ell}(\theta) e^{-in\psi}$$

with

$$d_{mn}^{\ell}(\theta) = \frac{\sin^{n-m}\theta (1+\cos\theta)^m}{2^{\ell} [(\ell+m)!(\ell-m)!]^{1/2}} \left[ \frac{(\ell-n)!}{(\ell+n)!} \right]^{1/2} \frac{d^{\ell+n}}{d(\cos\theta)^{\ell+n}} (\cos\theta - 1)^{\ell+m} (1 + \cos\theta)^{\ell-m}$$

where  $m$  and  $n$  are integers such that  $-\ell \leq m, n \leq \ell$ . The functions

$D_{mn}^{\ell}$  form a complete orthogonal system in the space of square integrable functions on  $SO(3)$  with respect to the inner product

$$\begin{aligned} \langle f_1, f_2 \rangle &= \int f_1(g) f_2(g) dg \\ &= \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} f_1(\phi, \theta, \psi) f_2(\phi, \theta, \psi) \sin \theta d\phi d\theta d\psi. \end{aligned}$$

An EFD(N) is a probability density on  $SO(3)$  of the form

$$p(\phi, \theta, \psi) = \exp \sum_{\ell=0}^N \sum_{m,n=-\ell}^{\ell} a_{mn}^{\ell} D_{mn}^{\ell}(\phi, \theta, \psi)$$

where  $a_{00}^0$  is a normalizing constant. By Theorem 10, any continuous probability density function can be approximated as closely as desired by such an EFD(N) in the aforementioned inner product space.

Let us now consider the following estimation problem: The signal process  $s_k$  is a sequence of random rotations on  $SO(3)$  which satisfies  $s_{k+1} = w_k \circ s_k$ ,  $\circ$  denoting product rotation, for some sequence of known rotations  $w_k$ . The measurement  $m_k$  is a concatenation of the signal  $s_k$  and the rotational white noise  $v_k$ , i.e.  $m_k = v_k \circ s_k$ . Suppose it is known that  $s_0$  and  $v_k$  have EFD(N)'s which are described by

$$p(s_0) = \exp \sum_{\ell=0}^N \sum_{m,n=-\ell}^{\ell} a_{mn}^{\ell 0} D_{mn}^{\ell}(s_0)$$

$$p(v_k) = \exp \sum_{\ell=0}^N \sum_{m,n=-\ell}^{\ell} b_{mn}^{\ell k} D_{mn}^{\ell}(v_k).$$

We would like to find the optimal estimate of  $s_k$  on  $SO(3)$  given the measurements  $m^k = \{m_1, \dots, m_k\}$  with respect to an error criterion which provides a measure of the deviation of the estimated orientation of the signal rotation  $s_k$  from the orientation of the signal rotation itself.

Following the calculation in the subsection V.3, we can show that the conditional density  $p(s_k | m^k)$  is an EFD(N) of the form

$$p(s_k | m^k) = \exp \sum_{\ell=0}^N \sum_{m,n=-\ell}^{\ell} a_{mn}^{\ell k} D_{mn}^{\ell}(s_k)$$

where the coefficients  $a_{mn}^{\ell k}$  are determined recursively by the formulas

$$a_{mn}^{\ell k} = \sum_{j=-\ell}^{\ell} \{a_{jn}^{\ell, k-1} D_{jm}^{\ell}(w_{k-1}^{-1}) + (-1)^{m+n} b_{j, -m}^{\ell k} D_{j, -n}^{\ell}(m_k)\}, \quad (32)$$

$$\ell \neq 0, k=1, 2, \dots$$

and  $a_{00}^{0k}$  is a normalizing constant. These formulas enable us to calculate the sequential conditional densities by updating recursively a finite and fixed number of parameters.

In order to define an error criterion for orientation estimation, it is necessary to have a measure of the distance between two orientations. We will first describe such a measure, using quaternions [41]. We recall that a rotation about an axis in the direction of a unit vector  $[\ell, m, n]$  through an angle  $\phi$  is represented by the (unit) quaternion

$$q = [q_1, q_2, q_3, q_4]' = [\cos \frac{\phi}{2}, \ell \sin \frac{\phi}{2}, m \sin \frac{\phi}{2}, n \sin \frac{\phi}{2}]'$$

Given two orientations, the minimal angle in radians required to bring one into the other is a natural measure of distance between them and defines a Riemannian metric on  $SO(3)$ . If the orientations are represented by the quaternions,  $q$  and  $p$ , and the minimal angle is denoted by  $\rho(q, p)$ , then we have  $q'p = \cos \frac{1}{2} \rho(q, p)$ . As  $(1 - \cos \rho)/2$  is a monotone increasing function of  $\rho$ , a measure of distance between  $p$  and  $q$  can be defined to be  $(1 - \cos \rho(q, p))/2 = 1 - (q'p)^2$ . It can be shown that if the orientations,  $q$  and  $p$ , are described by the  $3 \times 3$ -dimensional orthogonal matrices,  $Q$  and  $P$ , then this measure of distance can also be expressed as  $(3 - \text{tr } PQ')/4$ .

We are now ready to define the error criterion for orientation estimation. Let  $q$  be a random quaternion and  $p$  its estimate. Then a measure of the estimation error is

$$J(q, p) = E(1 - (q'p)^2).$$

If the probability distribution of  $q$  is given, the estimate  $p$  which minimizes  $J$  may be obtained from observing that

$$J(q, p) = 1 - p' E(qq') p.$$

It is well known that the quadratic form  $p'Vp$  of the positive definite matrix  $V = E(qq')$  is maximized over unit vectors  $p$  when  $p$  is an eigenvector associated with the largest eigenvalue  $\lambda$  of  $V$ . Moreover, the maximum value is  $\lambda$ .

Hence,

$$\begin{aligned} \min_p J(q, p) &= 1 - \hat{q}' E(qq') \hat{q} \\ &= 1 - \lambda \end{aligned}$$

where



$\lambda$  = the maximum eigenvalue of  $E(qq')$

$\hat{q}$  = the unit eigenvector of  $E(qq')$  associated with  $\lambda$ .

Using the conditional density  $p(s_k | m^k)$  that is computed recursively through (32), the optimal estimate of the orientation can then be determined as follows. First compute the conditional covariance matrix  $E(q(k)q'(k) | m^k)$  where  $q(k)$  is the quaternion for  $s_k$  whose components expressed in terms of the Euler angles are given below:

$$\begin{aligned} q_1 &= \cos \frac{\theta}{2} \cos \frac{\phi+\psi}{2} \\ q_2 &= \sin \frac{\theta}{2} \cos \frac{\phi-\psi}{2} \\ q_3 &= \sin \frac{\theta}{2} \sin \frac{\phi-\psi}{2} \\ q_4 &= \cos \frac{\theta}{2} \sin \frac{\phi+\psi}{2} \end{aligned} \quad (33)$$

Then use some standard numerical method to compute the largest eigenvalue  $\lambda(k)$  and the associated unit eigenvector  $\hat{q}(k|k)$ . The Euler angles  $(\hat{\phi}, \hat{\theta}, \hat{\psi})$  of the optimal estimate may then be determined from  $\hat{q}(k|k)$  through the equations,

$$\begin{aligned} \cos \hat{\theta} &= 2(\hat{q}_1^2 + \hat{q}_4^2) - 1, & 0 \leq \hat{\theta} \leq \pi \\ \sin \hat{\phi} &= -\frac{1}{\Delta} (\hat{q}_3 \hat{q}_1 + \hat{q}_2 \hat{q}_4), & \cos \hat{\phi} = \frac{1}{\Delta} (\hat{q}_1 \hat{q}_2 - \hat{q}_3 \hat{q}_4) \\ \sin \hat{\psi} &= \frac{1}{\Delta} (\hat{q}_2 \hat{q}_4 - \hat{q}_1 \hat{q}_3), & \cos \hat{\psi} = \frac{1}{\Delta} (\hat{q}_1 \hat{q}_2 + \hat{q}_3 \hat{q}_4) \end{aligned}$$

with

$$\Delta = \sqrt{(\hat{q}_1^2 + \hat{q}_4^2)(\hat{q}_2^2 + \hat{q}_3^2)}.$$

This simply inverts set of relationships (33). The estimation error is  $1-\lambda(k)$ .

## VI. DETECTION FOR CONTINUOUS-TIME SYSTEMS ON LIE GROUPS.

The idea of "rolling without slipping" introduced in Section IV will now be generalized and used to formulate an observation process on an arbitrary matrix Lie group. Briefly, we will inject the differentials of an observation process described by a vector Ito differential equation into a Lie group via the exponential map and then piece them together. The resulting product integral describes our observation process on the Lie group. The injected vector observation process is called its skew form.

The observation process thus constructed on a Lie group will be seen to satisfy a bilinear matrix stochastic differential equation, when its skew form is linear. The observational noise can be viewed as entering multiplicatively.

Given an arbitrary bilinear matrix observation process, we will show that the corresponding skew observation process can be obtained by "reversing" the above injecting procedure. Further, these two procedures will be seen to induce two "almost sure" bijective mappings between a vector-valued and a matrix-valued function space, one being the inverse of the other.

It is well known that the study of a Lie group may be greatly simplified by considering the tangent space (the Lie algebra) of the Lie group at its identity. In fact, the local study of a Lie group is entirely equivalent to the study of the algebraic structure of the Lie algebra. In this paper, the above bijective mappings facilitate similar simplification. It enables us to evaluate the likelihood ratio

in a finite dimensional linear space--the Lie algebra!

In view of the above construction, the null and the alternative hypotheses that the signal is respectively absent and present in the observation on a Lie group can be written in terms of a pair of bilinear matrix stochastic differential equations. Using the bijective mappings, we may transform these hypotheses on a Lie group into those on the corresponding Lie algebra. There the likelihood ratio can be expressed by the well-known formula in [43] and [44]. Thus the likelihood ratio on a Lie group can also be evaluated through least-squares estimation.

When the signal is a linear diffusion process, the idea of using the bijective mappings to work in the Lie algebra also leads to a finite dimensional filtering equation for evaluating the least-squares estimate. This equation is indeed an immediate extension of the Kalman-Bucy filter to the case of observation on Lie groups.



# VI.1 Almost Sure Representation of Continuous Curves on Lie Groups.

Let  $R^{n \times n}$  denote the set of real  $n \times n$  matrices and  $\{R_1, \dots, R_m\}$  a basis of a Lie algebra  $L$  in  $R^{n \times n}$ . Then the set

$$G = \{M : M = \exp(A_1) \exp(A_2) \dots \exp(A_k) ; A_i \in L,$$

$$i = 1, \dots, k ; k = 0, 1, \dots \}$$

is an  $m$ -dimensional Lie group related to  $L$  by a one-to-one map  $\phi_M$  from a neighborhood of  $0 \in L$  onto a neighborhood of  $M \in G$ . The map is defined by

$$\phi_M(A) = \exp(A)M, \quad A \in L.$$

A continuous curve in  $G$  is usually represented by an  $m \times m$ -matrix-valued continuous function on a closed interval  $T = [0, s]$  of the real line. In this section we will show that, under certain assumptions, a continuous curve in  $G$  starting from the identity element  $I \in G$  can also be represented by an  $m$ -vector-valued function on  $T$  in a certain "almost sure" sense.

We will use the following notation: :

$C_\ell$  = the family of continuous  $m$ -vector-valued functions,  $a$ , on  $T$  with initial value  $a(0) = 0$ ,

$C_g$  = the family of continuous  $m \times m$ -matrix-valued functions,  $A$ , on  $T$  such that  $A(t)$  is in  $G$  for each  $t$  in  $T$  and with initial value  $A(0) = I$ ,

$B_\ell$  = the Borel  $\sigma$ -field of  $C_\ell$  in the uniform topology,

$\mathcal{B}_g$  = the Borel  $\sigma$ -field of  $C_g$  in the uniform topology,

$w$  = a standard  $p$ -vector Brownian motion on a probability space,

$(\Omega, \mathcal{A}, P)$ ,

$$\left( \sum_{i=1}^m, \sum_{j=1}^m \right) \\ [ ] = \text{"the integral part of"}$$

Lower case letters will denote elements in  $C_\ell$  and upper case letters will denote elements in  $C_g$ .

Let  $y$  be an  $m$ -vector stochastic process on  $T$  satisfying the following Ito differential equation:

$$dy(t) = f(t)dt + Q^{\frac{1}{2}}(t)dw(t) \quad (34)$$

where  $f$  is an  $m$ -vector stochastic process on  $T$  and  $Q^{\frac{1}{2}} : T \rightarrow R^{m \times p}$  is Borel-measurable and bounded, i.e.,

$$\|Q^{\frac{1}{2}}(t)\|^2 \triangleq \text{tr } Q^{\frac{1}{2}}(t)(Q^{\frac{1}{2}}(t))^* \leq c_1^2, \text{ for } t \in T. \quad (35)$$

Let  $H_n : C_\ell \rightarrow C_g$  be defined by

$$\begin{aligned} (H_n(a))(t) &= I & (t = 0) \\ &= \exp \left[ \sum_j (a_j(t) - a_j(\ell 2^{-n})) R_j \right] (H_n(a))(\ell 2^{-n}) \\ & & (t \geq 0, \ell = [2^n t]) \end{aligned} \quad (36)$$

for  $a = [a_1, a_2, \dots, a_m] \in C_\ell$ .

Let  $K(\Delta) \triangleq \sum_j y_j(\Delta) R_j \triangleq \sum_j (y_j(t) - y_j(\ell 2^{-n})) R_j$  and  $Y_n(t) \triangleq (H_n(y))(t)$ .

Then

$$Y_n(t) - Y_n(\ell 2^{-n}) = (\exp(K(\Delta)) - I) Y_n(\ell 2^{-n})$$

Recall the following oscillation property ([45], p. 57),

$$P \left[ \lim_{\substack{t=t_2-t_1+0 \\ 0 \leq t_1 < t_2 \leq s}} \frac{|y(t_2) - y(t_1)|}{(2t \lg(1/t))^{\frac{1}{2}}} = \max_{t \leq T} \|Q^{\frac{1}{2}}(t)\| \right] = 1. \quad (37)$$

It is clear that up to terms involving  $K^3(\Delta)$  (of magnitude  $< 2^{-4n/3}$ ),

$$Y_n(t) - Y_n(\ell 2^{-n}) = (K(\Delta) + \frac{1}{2} K^2(\Delta)) Y_n(\ell 2^{-n}).$$

By simple calculations,

$$K^2(\Delta) \equiv \sum_i \sum_j Q_{ij}(t) R_i R_j \Delta t$$

Thus the definition of the Ito integral leads at once to the conjecture that  $Y \stackrel{\Delta}{=} \lim_{n \rightarrow \infty} Y_n$  is the solution of

$$dY(t) = \left[ \sum_j R_j dy_j(t) + M(t) dt \right] Y(t) \quad (38)$$

$$M(t) = \frac{1}{2} \sum_k \sum_j Q_{ij}(t) R_i R_j \quad (39)$$

$$Y(0) = I.$$

It is appropriate to remark here that the sequence of operators  $H_n$  was first devised in [46] to construct Brownian motion on the three dimensional rotation group,  $SO(3)$ , and later used in [45] to construct Brownian motion on a Lie group. Exactly the same trick was used to formulate an observation process on  $SO(3)$  in [47]. In this paper, this trick together with some techniques developed in [47] will be used, with little modification, to treat a large class of detection problems on arbitrary matrix Lie groups.

Following closely the six steps taken in Section 4.8 of [45] and keeping in mind the assumption (35) and the oscillation property (37), we come to the following conclusions:



(i) There is one and only one solution to (38)

(ii) The solution can be expressed as an almost surely convergent series as follows:

$$Y = \sum_{n \geq 0} \Theta_n$$

$$\Theta_n(t) = \int_0^t \left( \sum_i R_i dy_i(\tau) + M(\tau) d\tau \right) \Theta_{n-1}(\tau)$$

$$\Theta_0(t) = I$$

(iii) The sequence  $\{Y_n\}$  converges uniformly on  $T$  to the solution  $Y$  of (38) almost surely. In other words,  $\{H_n(a)\}$  converges uniformly on  $T$  to a continuous function  $H(a) \in C_g$  for each element  $a$  of a  $B_\ell$ -measurable set  $B_1 \subset C_\ell$  such that  $\mu_y(B_1) = 1$ , where  $\mu_y$  denotes the measure on  $(C_\ell, B_\ell)$  induced by  $y$ .

The operator  $H = \lim_{n \rightarrow \infty} H_n$  is the so-called product integral operator, which is usually used to solve matrix differential equations (see, e.g., [48] and [49]). Its application here to construct random processes on a Lie group by the use of random processes on its Lie algebra yields a random matrix differential equation (38), which is a global representation of the constructed random process on the Lie group rather than usual local representations for random processes on differential manifolds and Lie groups (see, e.g., [50] and [51]). This feature is obviously important in order to draw useful results for engineering purposes, as a sample path of a diffusion process may possibly zigzag across the boundary of a coordinate neighborhood infinitely frequently over a fixed time interval ([13], [14]).

We finally remark that the nonlinear diffusion  $x$  which satisfies

$$dx = g(x,t)dt + \sigma(x,t)dw$$

falls into the range of random processes  $y$  defined in (34).

In the following we will consider the converse problem of inducing a random process on a vector space by a random process on a Lie group. More specifically, we will construct the inverse operator,  $J$ , of  $H$  by defining the appropriate "inverse" operator  $J_n$  of  $H_n$ .

Let  $C_m$  denote the family of  $m \times m$ -matrix-valued continuous functions which are representations of continuous curves on  $L$ .

Let  $A \in C_g$  and  $n_1$  be the smallest integer such that for all  $n \geq n_1$  and  $0 \leq i \leq [s2^n]$ ,

$$||A((i+1)2^{-n})A^{-1}(i2^{-n}) - I|| \leq 1$$

Define  $K_n : C_g \rightarrow C_m$  by

$$(K_n(A))(t) = 0, \quad (t \in T), \quad (40)$$

for  $n < n_1$ , and

$$(K_n(A))(t) = 0, \quad (t = 0), \quad (41)$$

$$= (K_n(A))(\ell 2^{-n}) + \lg(A(t)A^{-1}(\ell 2^{-n}))$$

$$(t \geq 0, \quad \ell = [t2^n])$$

for  $n \geq n_1$ .

Setting  $(K_n(A))(t) = \sum_j R_j [(K_n(A))(t)]_j$ , we define

$J_n : C_g \rightarrow C_\ell$  by

$$(J_n(A))(t) = [(K_n(A))(t)]_1, \dots, [(K_n(A))(t)]_m]^t, \quad (42)$$

for  $A \in C_g$ . We will now show that  $\{J_n(A)\}$  converges uniformly on  $T$  to a continuous function  $J(A) \in C_\ell$ , for almost all  $A$  with respect to the measure  $\nu_Y$  on  $(C_g, B_g)$  induced by  $Y$ , constructed previously. Let



$$\Omega_n^1 = \{\omega \in \Omega \mid \|Y(i2^{-n}, \omega)Y^{-1}((i-1)2^{-n}, \omega) - I\| < 1\}$$

$$\Omega_n^t = \{\omega \in \Omega \mid \|Y(t, \omega)Y^{-1}(\ell 2^{-n}, \omega) - I\| < 1\}$$

$$\Omega_n = \bigcap_{i=1}^{\ell} \Omega_n^1 \cap \Omega_n^t.$$

Recalling (37) and (38) it can be easily seen that

$$\lim_{n \rightarrow \infty} P(\Omega_n) = 1.$$

For notational simplicity, we will denote  $(K_n(Y))(t)$  and

$$(K(Y))(t) \triangleq -\frac{1}{2} \int_0^t M(s) ds + \int_0^t (dY(s))Y^{-1}(s)$$

by  $K_n(t)$  and  $K(t)$ , respectively. Let

$$\begin{aligned} \bar{K}_n(t) \triangleq & \sum_{i=1}^{\ell} \left[ -\frac{1}{2} M((i-1)2^{-n})2^{-n} \right. \\ & + (Y(i2^{-n}) - Y((i-1)2^{-n})) \cdot Y^{-1}((i-1)2^{-n}) \\ & \left. + \left[ -\frac{1}{2} M(\ell 2^{-n})(t - \ell 2^{-n}) + (Y(t) - Y(\ell 2^{-n}))Y^{-1}(\ell 2^{-n}) \right] \right]. \quad (43) \end{aligned}$$

Then

$$\begin{aligned} E \|K_n(t) - K(t)\|^2 &= E[\text{tr}[(K_n(t) - K(t))(K_n(t) - K(t))']] \\ &\leq E \|K_n(t) - \bar{K}_n(t)\|^2 + E \|\bar{K}_n(t) - K(t)\|^2 \end{aligned} \quad (44)$$

Note that  $K_n(t) = 0$  on  $\Omega - \Omega_n$  by (40) and the definition of  $\Omega_n$ . Hence

$$E \|K_n(t) - \bar{K}_n(t)\|^2 = \int_{\Omega_n} \|K_n(t) - \bar{K}_n(t)\|^2 dP + \int_{\Omega - \Omega_n} \|\bar{K}_n(t)\|^2 dP. \quad (45)$$

By the definition of the Ito integral

$$\lim_{n \rightarrow \infty} E \|\bar{K}_n(t) - K(t)\|^2 = 0 \quad (46)$$

It is easy to see

$$\lim_{n \rightarrow \infty} \int_{\Omega - \Omega_n} \|\bar{K}_n(t)\|^2 dP \leq \lim_{n \rightarrow \infty} \|\bar{K}_n(t) - K(t)\|^2 dP + \lim_{n \rightarrow \infty} \int_{\Omega - \Omega_n} \|K(t)\|^2 dP \quad (47)$$

where the first term on the right hand side vanishes because of (46) and the second term vanishes because  $\lim_{n \rightarrow \infty} P(\Omega - \Omega_n) = 0$ . Substituting (41) and (43) into the first term on the right hand side of (45) yields

$$\begin{aligned} \int_{\Omega_n} \|K_n(t) - \bar{K}_n(t)\|^2 dP &\leq \int_{\Omega_n} \left\| -\frac{1}{2} \sum_{i=1}^{\ell} M((i-1)2^{-n}) (i2^{-n} - (i-1)2^{-n}) \right. \\ &\quad \left. + \frac{1}{2} \sum_{i=1}^{\ell} (Y(i2^{-n})Y^{-1}((i-1)2^{-n}) - I)^2 \right. \\ &\quad \left. - \frac{1}{2} M(\ell 2^{-n})(t - \ell 2^{-n}) + \frac{1}{2} (Y(t)Y^{-1}(\ell 2^{-n}) - I)^2 \right\|^2 dP \\ &+ \int_{\Omega_n} \left\| \sum_{k=1}^{\ell} \sum_{i \geq 3} [((-1)^{i-1}/i) (Y(k2^{-n})Y^{-1}((k-1)2^{-n}) - I)^i \right. \\ &\quad \left. + \sum_{i \geq 3} [((-1)^{i-1}/i) (Y(t)Y^{-1}(\ell 2^{-n}) - I)^i] \right\|^2 dP. \end{aligned}$$

With a view to (38), it can be easily proved that

$$\lim_{n \rightarrow \infty} \int_{\Omega_n} \|K_n(t) - \bar{K}_n(t)\|^2 dP = 0 \quad (48)$$

Combining (44) - (48), completes the proof that

$(K_n(Y))(t)$  converges to  $(K(Y))(t)$  in quadratic mean. Hence there is a subsequence  $\{n'\}$  of  $\{n\}$  such that with probability one

$$\lim_{n \rightarrow \infty} (K_{n'}(Y))(t) = (K(Y))(t) = -\frac{1}{2} \int_0^t M(s)ds + \int_0^t (dY(s))Y^{-1}(s). \quad (49)$$

Then it is easy to see from (42) that  $\{J_{n'}(A)\}$  converges uniformly on  $T$  to

$$J(A) = [[(K(A))(t)]_1, \dots, [(K(A))(t)]_m]' \in C_\ell$$

for each element  $A$  of a  $B_g$ -measurable set  $B_2 \subset C_g$  such that  $v_y(B_2) = 1$ . Furthermore,  $J$  is injective on  $B_2$  because the differential equation (38) with  $Y$  viewed as the unknown function has a unique solution so that  $K(A_1) = K(A_2)$  implies that  $A_1 = A_2$ , in view of (49). Comparing (49) with (38), we see that for each  $a \in B_1$ ,  $H(a)$  is an element of  $B_2 (B_1 \subseteq J(B_2))$  and  $J(H(a)) = a$  ( $J = H^{-1}$ ). We now come to the conclusion that almost all continuous curves in  $G$  with respect to  $v_y$  can also be represented by continuous  $m$ -vector-valued functions on  $T$ . Summarizing what has been shown, we obtain the following theorem.

Theorem 1. Let  $y$  be the solution of (34) and let  $H_n: C_\ell \rightarrow C_g$  and  $J_n: C_g \rightarrow C_\ell$  be defined by (36) and (42). Then  $\{H_n(y)\}$  converges to  $Y=H(y)$  with probability one, and  $Y$  satisfies

$$dY(t) = [\sum_i R_i dy_i(t) + 1/2 \sum_i \sum_j Q_{ij}(t) R_i R_j dt] Y(t)$$

with initial value  $Y(0)=I$ . Conversely,  $\{J_n(Y)\}$  converges to  $y=J(Y)$  with probability one, and  $y$  satisfies the above equation too. In fact,  $J=H^{-1}$  almost surely.

## VI.2. Hypotheses on Lie Groups and Evaluation of Likelihood Ratios.

The (almost sure) bijection constructed in the previous section will be used to formulate a detection problem on a matrix Lie group  $G$  and to derive a likelihood ratio formula as a function of the updated observation. Let us first write down a pair of hypotheses on the Lie algebra  $L$  of  $G$  in the form of  $m$ -vector Ito differential equations:



$$H_{11} : dy(t) = m(t)dt + Q^{\frac{1}{2}}(t)dw(t) \quad (50)$$

$$H_{00} : dy(t) = Q^{\frac{1}{2}}(t)dw(t) \quad (51)$$

where  $Q^{\frac{1}{2}}: T \rightarrow R^{m \times n}$  is Borel-measurable and bounded,

$$\|Q^{\frac{1}{2}}(t)\|^2 \triangleq \text{tr } Q^{\frac{1}{2}}(t)(Q^{\frac{1}{2}}(t))' \leq c_1, \text{ for } t \in T.$$

The process  $y$  is viewed as the skew form of the observation process and it is injected into  $G$  via  $H$ . Applying Theorem 1 and letting  $Y = H(y)$ , we obtain the following two hypotheses  $H_1$  and  $H_0$  on  $G$  corresponding to  $H_{11}$  and  $H_{00}$ , respectively.

$$H_1 : dY(t) = \left[ \sum_{j=1}^m R_j m_j(t)dt + \sum_{j=1}^n B_j(t)dw_j(t) + M(t)dt \right] Y(t) \quad (52)$$

$$H_0 : dY(t) = \left[ \sum_{j=1}^n B_j(t)dw_j(t) + M(t)dt \right] Y(t) \quad (53)$$

where

$$Y(0) = I$$

$$B_j(t) = \sum_{i=1}^m R_i Q_{ij}^{\frac{1}{2}}(t)$$

$$M(t) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m Q_{ij}(t) R_i R_j.$$

Let the measures on  $(C_g, B_g)$  induced by  $Y$  under  $H_1$  and  $H_0$  be denoted by  $\nu_1$  and  $\nu_0$  respectively. From the measure-theoretic viewpoint, the detection problem is to evaluate the Radon-Nikodym derivative  $d\nu_1/d\nu_0$  in  $(C_g, B_g)$ , assuming it exists.

Let the measures on  $(C_\ell, B_\ell)$  induced by  $y$  under  $H_{11}$  and  $H_{00}$  be denoted by  $\mu_1$  and  $\mu_0$  respectively. We note ([43], [44]) that if  $\int_0^s m'Q^{-1}mdt < \infty$ , a.s., (P), then

$$\frac{d\mu_1}{d\mu_0}(y) = \exp\left\{-\frac{1}{2} \int_0^s \hat{m}'(t)Q^{-1}(t)\hat{m}(t)dt + \int_0^s \hat{m}'(t)Q^{-1}(t)dy(t)\right\} \text{ a.s. (P)} \quad (54)$$

where  $\hat{m}(t) = E(m(t)|y^t, H_{11})$ ,  $y^t$  is the restrictions of  $y$  to  $[0, t]$ .

As  $Y = H(y)$  and  $H$  is almost surely bijective, we may anticipate that  $d\nu_1/d\nu_0$  is equal to  $d\mu_1/d\mu_0$  after some change of variables. This is indeed the case.

Lemma 5. Let  $\theta_1$  and  $\theta_2$  be any random objects taking values in the same measurable space  $(\Theta, \Theta_*)$ ,  $\Theta_*$  being a  $\sigma$ -field in  $\Theta$ . Let  $f$  be a measurable mapping from  $(\Theta, \Theta_*)$  into another measurable space  $(\Lambda, \Lambda_*)$  and  $\lambda_1 = f(\theta_1)$  and  $\lambda_2 = f(\theta_2)$ . Let the measures on  $(\Theta, \Theta_*)$  and  $(\Lambda, \Lambda_*)$  induced by  $\theta_1$  and  $\lambda_1$  be denoted by  $\xi_1$  and  $\eta_1$ , respectively, for  $i = 1$  and  $2$ . If  $\xi_1 \ll \xi_2$ , then  $\eta_1 \ll \eta_2$  and

$$\frac{d\eta_1}{d\eta_2}(\lambda) = E\left(\frac{d\xi_1}{d\xi_2} \mid \sigma(f)\right)(\lambda), \text{ a.s. } (\eta_2)$$

where  $\sigma(f)$  denotes the  $\sigma$ -subfield of  $\Theta_*$  generated by  $f$ .

If, in addition,  $f$  is bijective a.s.  $(\xi_2)$ , then

$$\frac{d\eta_1}{d\eta_2}(\lambda) = \frac{d\xi_1}{d\xi_2}(f^{-1}(\lambda)), \text{ a.s. } (\eta_2)$$

This lemma is an immediate extension of Lemma 3, p. 99 in [52].

A proof can be found in [47].

In view of this lemma, it is now easily seen that  $(dv_1/dv_0)(Y)$  is equal to the right side of (54) with  $y^t$  replaced by  $H^{-1}(Y^t) = J(Y^t)$ ,  $Y^t$  being the restriction of  $Y$  to  $[0, t]$ .

Let  $e_{ij}$  be the  $m \times m$ -matrix of which the  $(i, j)$  component is one and the other components are zero. Let  $\{R_{m+j}, j = 1, \dots, g^2 - m\}$  be  $g \times g$ -matrices such that  $\{R_j, j = 1, \dots, g^2\}$  form a basis of  $R^{g \times g}$ . Now we may write  $e_{ij} = \sum_{k=1}^{g^2} R_k e_{ij}^k$ , for some constants  $\{e_{ij}^k\}$ .

Let  $e_k = [e_{ij}^k]$  be the matrix of which the  $(i, j)$  component is  $e_{ij}^k$  (55)

Since  $\left[ \int_0^t (dY(\tau))Y^{-1}(\tau) - \int_0^t M(\tau)d\tau \right]$  belongs to  $L$  which is spanned by

$\{R_1, \dots, R_m\}$ , we have

$$\text{tr}\{e_j' \left[ \int_0^t (dY(\tau))Y^{-1}(\tau) - \int_0^t M(\tau)d\tau \right]\} = 0, \text{ for } j > m.$$

From (38), it can be shown by simple calculation that

$$\begin{aligned} y(t) = (J(Y))(t) &= [\text{tr}\{e_1' \left[ \int_0^t (dY(\tau))Y^{-1}(\tau) - \int_0^t M(\tau)d\tau \right]\}, \\ &\dots, \text{tr}\{e_m' \left[ \int_0^t (dY(\tau))Y^{-1}(\tau) - \int_0^t M(\tau)d\tau \right]\}]' \end{aligned} \quad (56)$$

where  $M$  is defined by (39).

Now we note that  $E(m(t)|J(Y^t), H_{11}) = E(m(t)|Y^t, H_1)$ , since  $J$  is bijective and  $Y^t$  and  $J(Y^t)$  generate the same  $\sigma$ -subfield of  $A$ .

Summarizing what has been shown, we obtain the following theorem.

Theorem 13. Given a matrix Lie group  $G$ , we can formulate a detection problem on it, which is described by the bilinear matrix Ito



equations, (52) and (52)<sup>1</sup>. If  $\int_0^s m' Q m dt < \infty$ , a.s. (P), the likelihood ratio can be written as

$$\frac{dv_1}{dv_0}(Y) = \exp \left\{ -\frac{1}{2} \int_0^s \hat{m}_t' Q^{-1}(t) \hat{m}_t dt + \int_0^s \hat{m}_t Q^{-1}(t) dy(t) \right\}$$

$$dy(t) = [\text{tr} \{ e_1' [(dY(t)) Y^{-1}(t) - M(t) dt] \}] , \dots ,$$

$$\text{tr} \{ e_m' [(dY(t)) Y^{-1}(t) - M(t) dt] \}]'$$

$$\hat{m}_t = E(m(t) | Y^t, H_1)$$

$$M(t) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m Q_{ij}(t) R_i R_j$$

and  $e_k$  is defined by (55).

### VI. 3. Detection for Bilinear Systems.

In the previous section we have shown how a detection problem on a matrix Lie group can be formulated if we are given the matrix Lie group. We treated a class of detection problems by starting with a  $C_\ell$  representation and injecting it to obtain a  $C_g$  representation. Motivated by the existence of the bijective operator discussed in subsection VI.1, we ask if we can reverse this process. I.e., can we start with a  $C_g$  representation and obtain a  $C_\ell$  representation? In this section we answer that question in detail and thereby treat a large class of bilinear detection problems.

In fact, in view of the bilinear form of the hypotheses  $H_1$  and  $H_0$ , a natural question arises as to the detection for problem arbitrary bilinear  $g \times g$ -matrix Ito differential equations of the following form:

$$H_1 : dY = \left[ \sum_{i=1}^{\alpha} A_i m_i dt + \sum_{i=1}^{\beta} B_i dw_i + \sum_{i=1}^{\gamma} C_i q_i dt \right] Y \quad (57)$$

$$H_0 : dY = \left[ \sum_{i=1}^{\beta} B_i dw_i + \sum_{i=1}^{\gamma} C_i q_i dt \right] Y \quad (58)$$

$$Y(0) = I$$

We assume:

- I. For simplicity in the illustration of our approach,  $A_1, B_1$  and  $C_1$  are constant matrices,
- II.  $\{A_i, i = 1, \dots, \alpha\}$  are linearly independent,
- III.  $w = [w_1, \dots, w_\beta]'$  is a standard  $\beta$ -vector Wiener process,
- IV.  $m = [m_1, \dots, m_\alpha]'$  is a measurable  $\alpha$ -vector stochastic process such that 
$$\sum_{i=1}^{\alpha} \int_0^S m_i^2(t) dt < \infty, \text{ a.s. (P) ,}$$
- V.  $q = [q_1, \dots, q_\gamma]'$  is a measurable  $\gamma$ -vector stochastic process such that 
$$\sum_{i=1}^{\gamma} \int_0^S q_i^2(t) dt < \infty, \text{ a.s. (P), and } q(t) \text{ is } Y^t\text{-measurable,}$$
- VI.  $w$  is independent of  $m$ .

Comparing these hypotheses with those defined by (52) and (53), we anticipate a solution similar to that in the previous section. We notice that if the solution  $Y$  to both (57) and (58) is on a matrix Lie group ( $Y(t)$  is then nonsingular), and  $m$  and  $w$  enter the observation via the Lie algebra then an almost sure bijection  $J$  exists which transforms  $Y$  into a causally equivalent vector process. The detection problem at hand is then readily solved.

Let  $L_n$  denote the Lie algebra generated by  $\{A_1, \dots, A_\alpha, B_1, \dots, B_\beta\}$ . Assume that  $L_n$  is an  $n$ -dimensional linear space of which  $\{R_1, \dots, R_n\}$  is a basis such that  $R_i = A_i$ , for  $i = 1, \dots, \alpha$ . Since  $B_1 \in L_n$  we may write

$$B_1 = \sum_{j=1}^n \theta_j R_j, \quad i = 1, \dots, \beta.$$

and  $\theta_j, j = 1, \dots, n$  are the coefficients of  $B_1$  in the basis



$\{R_1, \dots, R_n\}$ . Let  $Q^{\frac{1}{2}}$  denote the  $n \times \beta$  matrix,  $[Q_{ij}^{\frac{1}{2}}]$ .

Assume that the Lie algebra  $L$  generated by  $\{A_1, \dots, A_\alpha, B_1, \dots, B_\beta, C_1, \dots, C_\gamma, R_i R_j, i = 1, \dots, n, j = 1, \dots, \beta\}$  is an  $m$ -dimensional linear space. In view of the results in Section 2, we may presume that a skew observation  $\underline{y} = J(Y)$  in the form of (50) exists and write

$$d\underline{y} = \underline{f} dt + Q^{\frac{1}{2}} dw, \quad (60)$$

where the  $m$ -vector stochastic process  $\underline{f}$  and the  $m \times \beta$ -matrix  $Q^{\frac{1}{2}}$  are to be determined.

Substituting (60) into (38) we obtain

$$\underline{f} = \underline{m} + \underline{r}$$

$$\underline{m} = [m_1, \dots, m_\alpha, 0, \dots, 0]'$$

$$\underline{r} = [r_1, \dots, r_m]'. \quad (61)$$

$$\sum_{i=1}^{\gamma} C_i q_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{\beta} Q_{ij} R_i R_j = \sum_{i=1}^m R_i r_i$$

( $Q_{ij}$  is the  $(i, j)$  component of the  $n \times n$ -matrix  $Q = Q^{\frac{1}{2}}(Q^{\frac{1}{2}})'$ ; in general,  $Q_{ij} \neq (Q_{ji}^{\frac{1}{2}})^2$ )

$$Q^{\frac{1}{2}} = \begin{bmatrix} Q^{\frac{1}{2}} \\ 0 \end{bmatrix} \begin{array}{c} \updownarrow n \\ \updownarrow m-n \\ - \end{array}$$

|  $\leftrightarrow$  |  
 $\beta$

Hence  $\underline{y} = H(Y)$  under  $H_1$  satisfies

$$d\underline{y} = \underline{g} dt + Q^{\frac{1}{2}} dw + \underline{z} dt$$

and similarly  $\underline{y} = J(Y)$  under  $H_0$  satisfies

$$H_{10} : d\underline{y} = Q^{\frac{1}{2}} dw + \underline{r} dt .$$

This shows that the hypotheses, (57) and (58), are interpretable on the Lie group  $G$  of  $L$ . Therefore,  $Y(t)$  is invertible and, by Theorem 1,  $\underline{y}$  and  $Y$  are causally equivalent.

Let  $C_g$  and  $B_g$  be defined as in the previous sections, and let the measures on  $(C_g, B_g)$  induced by  $Y$  under  $H_1$  and  $H_0$  be denoted by  $\mu_1$  and  $\mu_0$  respectively. In the following, we will evaluate the Radon-Nikodym derivative  $d\mu_1/d\mu_0$ , assuming it exists.

Let  $C^k$  ( $k=\alpha$  or  $k=m-\alpha$ ) denote the family of continuous  $k \times k$ -matrix-valued functions,  $A$ , on  $T$  with initial value  $A(0) = I$ , and let  $B^k$  denote the Borel  $\sigma$ -field of  $C^k$ .

Let  $y = [y_1, \dots, y_\alpha]'$  and let  $v_1^1$  be the measure on  $(C^\alpha, B^\alpha)$  induced by  $y$  under  $H_1$ . Let  $z = [y_{\alpha+1}, \dots, y_m]'$  and let  $v_1^2$  be the measure on  $(C^{m-\alpha}, B^{m-\alpha})$  induced by  $z$  under  $H_1$ . Then the measure on  $(C^m, B^m)$  induced by  $\underline{y}$  under  $H_1$  is equal to the product measure  $v_1^1 \times v_1^2$ . It is easy to see that  $v_1^2 = v_0^2$  and thus  $\frac{dv_1^2}{dv_0^2} = 1$ . Using a well-known lemma

(Lemma 2, p. 99 in [52]), we have, for  $t \in T$ ,

$$\frac{d(v_1^1 \times v_1^2)}{d(v_0^1 \times v_0^2)}(\underline{y}^t) = \frac{dv_1^1}{dv_0^1}(y^t) \frac{dv_1^2}{dv_0^2}(z^t) = \frac{dv_1^1}{dv_0^1}(y^t), \quad (62)$$

provided  $\frac{dv_1^1}{dv_0^1}$  exists.

We let  $I_1^1$  be the  $(\alpha \times \alpha)$  matrix with 1's on the main diagonal and 0's elsewhere.

We note that  $y$  under  $H_1$  satisfies

$$H_{21} : dy = mdt + I_n^\alpha Q^{\frac{1}{2}} dw + rdt$$

while  $y$  under  $H_0$  satisfies

$$H_{20} : dy = I_n^\alpha Q^{\frac{1}{2}} dw + rdt ,$$

where  $r = I_m^\alpha \underline{r}$ . If  $\det(I_n^\alpha Q I_n^\alpha) \neq 0$ , it is known that

the likelihood ratio of  $H_{21}$  to  $H_{20}$  can be written as:

$$\begin{aligned} \frac{dv_1^1}{dv_0^1}(y) = \exp[ & -\frac{1}{2} \int_0^S \hat{m}'(t) (I_m^\alpha Q I_m^\alpha)^{-1} \hat{m}(t) dt - \int_0^S \hat{m}(t) (I_m^\alpha Q I_m^\alpha)^{-1} r(t) dt \\ & + \int_0^t \hat{m}'(t) (I_m^\alpha Q I_m^\alpha)^{-1} dy(t)] \end{aligned} \quad (63)$$

where  $\hat{m}(t) = E(m(t) | y^t, H_{21})$ . We note that the assumption that  $q(t)$  is  $Y^t$ -measurable is used

Since  $J$  is bijective, we have, by Lemma 5,

$$\frac{d\mu_1}{d\mu_0}(Y) = \frac{d(v_1^1 \times v_1^2)}{d(v_0^1 \times v_0^2)}(J(Y)) = \frac{dv_1^1}{dv_0^1}(J(Y)) \quad (64)$$

Let  $e_k$  be defined as in (55) and write  $J(Y)$  as (56).

Substituting (63) into (64) then leads to the following theorem.

Theorem 14. Consider the two hypotheses  $H_1$  and  $H_0$  described by (57) and (58). Let the constant matrix  $Q$  be defined by (59) and assume that  $\det(I_n^\alpha Q I_n^\alpha) \neq 0$ . Then the likelihood ratio for  $H_1$  against  $H_0$  can be expressed, given a realization of  $Y$ , as follows:



$$\frac{dv_1}{dv_0}(Y^t) = \exp[-1/2 \int_0^t \hat{m}_s' (I_m^\alpha Q I_m^\alpha)^{-1} \hat{m}_s ds - \int_0^t \hat{m}_s' (I_m^\alpha Q I_m^\alpha)^{-1} r(s) ds + \int_0^t \hat{m}_s' (I_m^\alpha Q I_m^\alpha)^{-1} dy(s)]$$

where

$$\hat{m}_s = E(m(s) | Y^s, H_1)$$

$$dy(s) = [\text{tr}(e_1' [(dY(s))Y^{-1}(s) - Mds]), \dots, \text{tr}(e_\alpha' [(dY(s))Y^{-1}(s) - Mds])]'$$

$$M = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^{\beta} Q_{ij} R_i R_j'$$

and  $I_m^\alpha$ ,  $e_k$ , and  $r(= \frac{1}{m} \alpha r)$  are defined by (62), (55) and (61),

respectively.

#### VI. 4. Least-Squares Estimation.

In view of Theorem 13 and Theorem 14, it is noted that the evaluation of the likelihood ratio depends on the evaluation of the conditional expectation  $\hat{m}_s = E(m(s) | Y^s, H_1)$ . We recall that under  $H_1$

$$dY = \left[ \sum_{i=1}^{\alpha} A_i m_i dt + \sum_{i=1}^{\beta} B_i dw_i + \sum_{i=1}^{\gamma} C_i q_i dt \right] Y \quad (65)$$

$$Y(0) = I.$$

The evaluation of  $\hat{m}_s$  is thus a nonlinear filtering problem.

In this section we will use the transformation techniques developed in the previous sections to solve this filtering problem under the assumption that  $w(t) = W(t)x(t)$ , for  $t \in T$ , (66)

$$dx(t) = F(t)x(t)dt + G(t)dv(t) \quad (67)$$

$$x(0) = x_0$$

Here  $x(t)$  is a random vector,  $H(t)$ ,  $F(t)$  and  $G(t)$  are matrices of appropriate dimension,  $v$  is a standard Wiener process,  $x_0$  is a normal random vector and  $w$ ,  $v$ ,  $x_0$  are statistically independent.

It has been seen in the previous section that

$$E(m(t) | Y^t, H_1) = E(m(t) | \underline{y}^t, H_{11}) = E(m(t) | y^t, H_{21}), \text{ a.s. } (P). \quad (68)$$

Let  $C^\alpha$  denote the family of  $\alpha$ -vector-valued continuous functions and  $B^\alpha$  its Borel  $\sigma$ -field in the uniform topology. Then there exist a  $B_g^\alpha$ -measurable functional  $f_1: C_g^\alpha \rightarrow R^\alpha$  and a  $B^\alpha$ -measurable functional  $f_3: C^\alpha \rightarrow R^\alpha$  such that

$$f_1(Y^t) = E(m(t) | Y^t, H_1), \text{ a.s.}$$

$$f_3(Y^t) = E(m(t) | y^t, H_{21}), \text{ a.s.}$$

and

$$f_1(Y^t) = f_3(I_n^\alpha H(Y^t)), \text{ a.s.} \quad (69)$$

where  $I_n^\alpha H: C_g^\alpha \rightarrow C^\alpha$  is defined by  $(I_n^\alpha H(Y))(t) = I_n^\alpha [(H(Y))(t)] = y(t)$ .

In the following we will denote  $f_1(Y^t)$  by  $\hat{m}_t$  and  $f_3(y^t)$  by  $\hat{m}(t)$ .

We note that this convention is consistent with the previous sections.

Under the assumptions (66) and (67) it is well known that  $\hat{m}(t)$  satisfies the following Kalman-Bucy filtering equations,

$$\hat{m}(t) = H(t)\hat{x}(t)$$

$$d\hat{x}(t) = F(t)\hat{x}(t)dt + I_n^\alpha r dt + P(t)H'(t)(I_m^\alpha Q I_m^\alpha)^{-1}(dy(t) - H(t)\hat{x}(t)dt)$$

$$P(t) = F(t)P(t) + P(t)F'(t) - P(t)H'(t)(I_m^\alpha Q I_m^\alpha)^{-1}H(t)P(t) + G(t)G'(t) \quad (70)$$

$$P(0) = E(x_0 x_0').$$

Using these equations together with (68) and (69), we obtain the following theorem.

Theorem 15. Let the message process  $m$  and the observation process  $Y$  be as described by the equations, (66), (67), and (65). Then the conditional expectation  $\hat{m}_t$  satisfies

$$\hat{m}_t = H(t)\hat{x}_t$$

$$d\hat{x}_t = F(t)\hat{x}_t dt + r dt + P(t)H'(t)(I_m^\alpha Q I_\alpha^m)^{-1}(dy(t) - H(t)\hat{x}_t dt)$$

$$dy(t) = [\text{tr}\{e_1'[(dY(t))Y^{-1}(t) - Mdt]\}, \dots,$$

$$\text{tr}\{e_\alpha'[(dY(t))Y^{-1}(t) - Mdt]\}]',$$

where  $P$ ,  $I_m^\alpha$ ,  $e_k$ ,  $r$ ,  $M$ ,  $Q$  are determined by (70), (62), (55), (61), (39), (59), respectively.



REFERENCES

1. D. Q. Mayne and R. W. Brockett, eds., Geometric Methods in System Theory, D. Reidel, Boston, Mass., 1973.
2. P. d'Alessandro, A. Isidori, and R. Ruberti, Lectures on Bilinear System Theory, Springer-Verlag, New York, 1972.
3. R. W. Brockett, "Nonlinear Systems and Differential Geometry," Proc. of the IEEE, vol. 64, No. 1, Jan. 1976.
4. K. V. Mardia, Statistics of Directional Data, Academic Press, New York, 1972.
5. J. T. H. Lo, "Exponential Fourier Densities and Optimal Estimation and Detection on the Circle," IEEE Trans. Inform. Theory, vol. IT-23, pp. 110-116, Jan. 1977.
6. J. T. H. Lo and A. S. Willsky, Estimation for Rotational Processes with One Degree of Freedom, Harvard Tech Rept. No. 635, Div. of Eng. and Appl. Phys., Harvard Univ., July 1972.
7. J. T. H. Lo and L. R. Eshleman, "Exponential Fourier Densities on  $SO(3)$  and Optimal Estimation and Detection for Rotational Processes," Proc. IEEE Conf. on Decision and Control, pp. 134-140, December 1976.
8. J. T. H. Lo, "Signal Detection for Bilinear Systems," Information Sciences 9, pp. 249-278, 1975.
9. J. T. H. Lo and L. R. Eshleman, "Exponential Fourier Densities on  $S^2$  and Optimal Estimation and Detection for Directional Processes," IEEE Trans. Inform. Theory, vol. IT-23, pp. 321-336, May 1977.

10. J. T. H. Lo and L. R. Eshleman, "Exponential Fourier Densities on  $S^2/\pm 1$  and Optimal Estimation for Axial Processes," UMBC Mathematics Department, Research Rep. No. 77-1, Jan. 1977.
11. R. S. Bucy, C. Hecht, and K. D. Senne, "New Methods for Nonlinear Filtering," Revue Francaise d'Automatique, Informatique et de Recherche Operationnelle, February 1973, pp. 3-54.
12. R. S. Bucy and H. Youssef, "Fourier Realization of the Optimal Phase Demodulator," Fourth Symposium on Nonlinear Estimation and Its Applications, San Diego, California, 1973.
13. R. S. Bucy and A. J. Mallinckrodt, "An Optimal Phase Demodulator," Stochastics, vol. 1, pp. 3-23, 1973.
14. A. S. Willsky, "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication, Parts I, II," IEEE Trans. on Inf. Theory, vol. IT-20, No. 5, September 1974, pp. 577-590.
15. S. I. Marcus, Estimation and Analysis of Nonlinear Stochastic Systems, Ph.D. Thesis, Department of Electrical Engineering, M.I.T., Cambridge, Mass., June 1975; also M.I.T. Electronic Systems Laboratory Report No. ESL-R-601.
16. A. S. Willsky and S. I. Marcus, "Estimation for Bilinear Stochastic Systems," Proc. Second U.S.-Italy Seminar, A. Ruberti and R. R. Mohler, eds., Springer-Verlag, 1974.
17. C. Johnson and E. B. Stear, "Filtering with Multiplicative Noise in Finite Dimensional Euclidean Space," Proc. Sixth Symposium on Nonlinear Estimation Theory and Its Applications, San Diego, California, 1975.

18. G. H. Hardy, J. E. Littlewood, G. Polya, Inequalities,  
Cambridge, 1952.
19. G. Szegő and G. Pólya, Isoperimetric Inequalities in  
Mathematical Physics, Princeton, 1951.
20. S. Sherman, "Non Mean-Square Error Criteria," IRE Trans. on  
Inf. Theory, vol. 4, 1958, pp. 125-126 .
21. S. Sherman, "A Theorem on Convex Sets with Applications,"  
Ann. Math. Stat., vol. 26, 1955, pp. 763-767.
22. R. Bellman, A Brief Introduction to Theta Functions, Holt,  
Rinehart and Winston, New York, 1961.
23. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis,  
Amer. Ed., The MacMillan Company, New York, 1945.
24. J. L. Brown, "Asymmetric Non-Mean-Square Error Criteria,"  
IRE Trans. on Aut. Cont., vol. 7, 1962, pp. 64-66.
25. J. T. H. Lo and A. S. Willsky, "Estimation for Rotational  
Processes with One Degree of Freedom, Part I: Introduction and  
Continuous-Time Processes, Part II: Discrete-Time Processes,  
Part III: Implementation," IEEE Trans. Automa. Control,  
vol. AC-20, pp. 10-33, 1975.
26. P. Langevin, "Magnetisme et theorie des electrons," Ann. de  
Chim. et de Phys., vol. 5, pp. 70-127, 1905.
27. R. Von Mises, "Über die 'anzzahligkeit' der Atomgewicht und  
Verwandte Fragen" Physikal. Z., vol. 19, pp. 490-500, 1918.
28. T. T. Kadota and L. A. Shepp, "Conditions for Absolute  
Continuity between a Certain Pair of Probability Measures,"  
Z. Wahrscheinlichkeitstheorie verw. Geb. 16, pp. 250-260, 1970.



29. J. T. Lo, "A General Bayes Rule and Its Application to Nonlinear Estimation," Information Sciences 8, pp. 189-198, 1975.
30. J. T. Lo, "On Optimal Nonlinear Estimation - Part I: Continuous Observation," Proc. of the 8th Annual Allerton Conference on Circuits and Systems, 1970; Information Sciences 6, pp. 19-32, 1973.
31. R. L. Stratonovich, "Conditional Markov Processes," Theory of Probability and Its Applications, vol. 5, 1962.
32. H. J. Kushner, "On the Dynamic Equations of Conditional Probability Density Functions, with Applications to Optimal Stochastic Control Theory," J. of Math. Anal. Appl., vol. 8, 1964.
33. T. Kailath and P. Frost, "An Innovations Approach to Least Squares Estimation - Part II: Linear Smoothing in Additive White Noise," IEEE Trans. Automatic Control, vol. AC-13, December 1968, pp. 655-660.
34. F. W. Warner, Foundations of Differentiable Manifolds and Lie Groups, Scott, Foresman and Company, Glenview, Ill., 1971.
35. J. D. Talman, Special Functions: A Group Theoretic Approach, Benjamin, New York, 1968.
36. E. Hewitt and K. A. Ross, Abstract Harmonic Analysis I, Academic Press, New York, 1963.
37. N. J. Vilenkin, Special Functions and The Theory of Group Representations, Translations of Mathematical Monographs, vol. 22, Amer. Math. Soc., Providence, Rhode Island, 1968.

38. M. A. Krasnoselskii, Topological Methods in the Theory of Nonlinear Integral Equations, MacMillan, New York, 1964.
39. F. D. Murnaghan, The Theory of Group Representations, Johns Hopkins Press, Baltimore, 1938.
40. E. P. Wigner, "Symmetry Principles in Old and New Physics," Bull. Amer. Math. Soc. 74, pp. 793-815, 1968.
41. P. DuVal, Homographies, Quaternions, and Rotations, Oxford Math. Monographs, Clarendon Press, Oxford, 1964.
42. E. Isaacson and H. B. Keller, Analysis of Numerical Methods, Wiley, New York, 1966.
43. T. E. Duncan, "Evaluation of Likelihood Functions," Information and Control, vol. 13, 1968.
44. T. Kailath, "A General Likelihood-Ratio Formula for Random Signals in Gaussian Noise," IEEE Trans. on Information Theory, vol. IT-15, No. 3, May 1969.
45. H. P. McKean, Jr., Stochastic Integrals, Academic Press, New York, 1969.
46. H. P. McKean, Jr., "Brownian Motions on the 3-Dimensional Rotation Group, Mem. Coll. Sci. Kyoto Univ. 33, pp. 25-28, 1960.
47. J. T. Lo, "Signal Detection of Rotational Processes and Frequency Demodulation," presented at the 7th Annual Princeton Conference on Information Sciences and Systems, March, 1973.
48. L. Schlesinger, "Weitere Beitrage aum Infinitesimalkalkul der Matrizen," Math. Z. 35, pp. 33-61, 1932.
49. L. Schlesinger, "Neue Grundlagen fur einen Infinitesimalkalkul der Matrizen," Math. Z. 33, pp. 33-61, 1931.

50. K. Ito, "On Stochastic Differential Equations on a Differentiable Manifold I," Nagoya Math. J. 1, pp. 35-47, 1950.
51. K. Ito, "Brownian Motions on a Lie Group," Proc. Japan Acad., vol. 26, pp. 4-10, 1950.
52. A. V. Skorokhod, Studies in Theory of Random Processes, Addison-Wesley, Reading, Massachusetts, 1965.
53. F. Perrin, "Etude Mathematique du Mouvement Brownien de Rotation," Ann. Ecole Norm. Sup., vol. 45, 1928, pp. 1-51.
54. Y. H. Ku, L. P. Sheporaitis, "Global Properties of Diffusion Processes on Cylindrical Type Phase Space," J. of Franklin Institute, vol. 289, No. 2, Feb. 1970, pp. 1087-1103.



AD-A053 633

MARYLAND UNIV BALTIMORE COUNTY BALTIMORE DIV OF MATH--ETC F/G 12/1  
ESTIMATION AND DETECTION OF LIE GROUPS.(U)

OCT 77 J T LO

AFOSR-74-2671

UNCLASSIFIED

MRR-77-11

AFOSR-TR-78-0734

NL

2 OF 2  
ADA  
053633



END  
DATE  
FILMED  
6-78  
DDC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER <b>AFOSR-TR-78-0734</b>	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) <b>ESTIMATION AND DETECTION ON LIE GROUPS.</b>		5. TYPE OF REPORT & PERIOD COVERED <b>Interim rept.</b>	
6. AUTHOR(s) <b>JAMES TING-HO LO</b>		8. CONTRACT OR GRANT NUMBER(s) <b>15 AFOSR-74-2671</b>	
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Maryland Baltimore County Department of Mathematics Baltimore, Maryland 21228		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>61102F 2304 A1</b>	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NM) Bolling AFB, Washington D.C. 20332		12. REPORT DATE <b>22 OCT 1977</b>	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES <b>93</b>	
		15. SECURITY CLASS. (of this report) <b>UNCLASSIFIED</b>	
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) <b>14 MRB-77-22</b>			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Exponential Fourier density, Optimal estimation, Conditional distribution, Error criteria, Rolling without slipping, Circle, Lie group, Detection, Likelihood ratio, Bilinear system, Additive noise, Noise on group			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A new approach developed in recent years for nonlinear estimation and detection is the employment of the differential geometric techniques. In this report, this approach is surveyed and the ideas of "rolling without slipping" and the exponential Fourier densities are discussed in great detail. Some new results on estimation on general compact Lie groups are also included.			

409 579